

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**This Page Blank (uspto)**

**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2002, 20:17:55 ; Search time 3156 Seconds  
(without alignments)  
3061.512 Million cell updates/sec

Title: US-09-848-852A-3

Perfect score: 1772

Sequence: 1 MYSAQRFWGTIARRGAHLA.....DMPKLRRIYKELCHCKLTV 332

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=fragnet\_p2n.model -DEV=xlp  
-Q=/cgn2.1/USTFO\_spool/US09848852/runat\_16122002\_132040\_19132/app\_query.fasta\_1.519  
-DB=GenEmbl -Qfmt=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09848852 -CGN 1.1 3637 @runat\_16122002\_132040\_19132 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1751	98.8	1782	9 HSM800198	AL050283 Homo sapi
2	1751	98.8	2206	6 AX081143	AX081143 Sequence
3	1751	98.8	2224	9 AY008763	AY008763 Homo sapi
4	1751	98.8	2258	9 AF199459	AF199459 Homo sapi
5	1751	98.8	2269	9 HSM805332	AL834294 Homo sapi
6	1696	95.7	1707	10 AF194031	AF194031 Mus muscu
7	1696	95.7	2174	10 AY008764	AY008764 Mus muscu
8	969.5	54.7	139405	2 AC126239	AC126239 Felis cat
9	880	49.7	161428	2 AC126925	AC126925 Canis fam
10	849.5	47.9	153553	2 AC126921	AC126921 Bos tauru
11	820	46.3	1578	9 AC000923	AC000923 Homo sapi
12	794	44.8	257644	2 AC099294	AC099294 Rattus no
13	791	44.6	2007	9 AB060892	AB060892 Macaca fa
14	789	44.5	1830	9 BC008589	BC008589 Homo sapi
15	789	44.5	2793	9 BC030705	BC030705 Homo sapi
16	788.5	44.5	2673	9 AB074445	AB074445 Macaca fa
17	781.5	44.1	164504	2 AC129071	AC129071 Pan trogl
18	781.5	44.1	177703	2 AC016876	AC016876 Homo sapi
19	781.5	44.1	215795	2 AC127470	AC127470 Pan trogl
20	768	43.3	173588	2 AC127468	AC127468 Papio cyn
21	746	42.1	138792	2 AC119115	AC119115 Rattus no
22	726.5	41.0	76754	2 AC097861	AC097861 Rattus no
23	719	40.6	177555	2 AC130192	AC130192 Sus scrof
24	711.5	40.2	203083	2 AC069459	AC069459 Mus muscu
25	711.5	40.2	234182	10 AL603707	AL603707 Mouse DNA
26	691	39.0	203281	2 AC126237	AC126237 Canis fam
27	674	38.0	184026	2 AC098923	AC098923 Rattus no
28	661.5	37.3	503	6 AX209870	AX209870 Sequence
29	658	37.1	251835	2 AC099436	AC099436 Rattus no
30	566.5	32.0	203300	2 AC125091	AC125091 Mus muscu
31	519	29.3	329337	2 AL806532	AL806532 Mus muscu
32	474.5	26.8	5318	6 AX069229	AX069229 Sequence
33	474.5	26.8	6732	9 AF175325	AF175325 Homo sapi
34	473.5	26.7	164504	2 AC129071	AC129071 Pan trogl
35	456	25.7	1932	9 AF149770	AF149770 Homo sapi
36	456	25.7	2512	6 AX081135	AX081135 Sequence
37	453.5	25.6	2903	10 BC023129	BC023129 Mus muscu
38	396.5	22.4	150136	10 AC122027	AC122027 Mus muscu
39	396.5	22.4	267305	2 AC125216	AC125216 Mus muscu
40	395	22.3	182718	2 AC103148	AC103148 Rattus no
41	393	22.2	2484	9 AK074357	AK074357 Homo sapi
42	393	22.2	2497	9 AF151697	AF151697 Homo sapi
43	393	22.2	2795	6 AX081141	AX081141 Sequence
44	393	22.2	3031	9 HSM805453	AL834380 Homo sapi
45	393	22.2	3133	6 BD012321	BD012321 Genes rel

ALIGNMENTS

RESULT 1

HSM800198 1782 bp mRNA linear PRI 10-MAR-2001  
 LOCUS Homo sapiens mRNA; cDNA DKF2p586K0919 (from clone DKF2p586K0919);  
 DEFINITION complete cds.  
 ACCESSION AL050283  
 VERSION AL050283.1 GI:4886466  
 KEYWORDS human.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1782)  
 Wiemann,S., Well,B., Wellenreuther,R., Gassenhuber,J., Glassl,S.,  
 Ansorge,W., Boecker,M., Bloecher,H., Bauersachs,S., Blum,H.,  
 Lauber,J., Duesterhoeft,A., Beyer,A., Koehler,K., Strack,N.,  
 Mewes,H.W., Ottenwaelder,B., Obermaier,B., Tampe,J., Heubner,D.,  
 Wambutt,R., Korn,B., Klein,M. and Poustka,A.  
 Toward a Catalog of Human Genes and Proteins: Sequencing and  
 Analysis of 500 Novel Complete Protein Coding Human cDNAs  
 Genome Res. 11 (3), 422-435 (2001)  
 11230166  
 2 (bases 1 to 1782)  
 Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.  
 Direct Submission  
 Submitted (10-MAR-1999) MIPS, Am Klopferspitz 18a, D-82152  
 Martinsried, GERMANY  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
 sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing  
 consortium of the German Genome Project.  
 This clone (DKF2p586K0919) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further  
 information about the clone and the sequencing project is available  
 at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

## FEATURES

Location/Qualifiers  
 1..1782  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DKF2p586K0919"  
 /tissue\_type="uterus"  
 /clone\_lib="586 (synonym: hutel). Vector pSport1; host  
 DH10B; sites NotI + SalI/MluI"  
 /dev\_stage="adult"  
 204..1319  
 /gene="DKF2p586K0919"  
 204..1319  
 /gene="DKF2p586K0919"  
 /note="similarity to S.pombe SPBC19G7.09"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="CAB43384.1"  
 /db\_xref="GI:4886467"  
 /db\_xref="SPTREMBL:Q9Y3W9"  
 /translation="MAEDGVRSPPVSGPPEEDGLRWTPKSLPDPDGLLSCTLPN  
 GFQSGGEGRSAPPDASILLISNVCSIGDHVAQELFGSDGLMAEAEPRGKAGQ  
 HSPLEERHEVQSTLDFLOTGLSLPLSTDEVVEKLEDFQOEFSPSRKGLVLQI  
 IQSVORPMGNVGRFVAYKRHLVTMDDLGLTYGONLNDVNMVYGDLYMDTVPEK  
 VHFNSFFYDKLRKGYGVKRWTKNVDIFNKKLLLIPIHLEVHSLISVDVRRRTIT  
 YFDSORTLNRRKPHIAKYLQAEAVKKRDLDFHOGWKGYFKMNVARQNNDSDCGAVFL  
 QYCKHLALSQPSFTQQDMPKLRRLQIYKELCHCLTV"  
 1741..1746  
 polyA\_signal  
 polyA\_site  
 BASE\_COUNT 419 a 466 c 475 g 422 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 8e-165 Length: 1782  
 Score: 1751.00 Matches: 331  
 Percent Similarity: 99.40% Conservative: 0  
 Best Local Similarity: 99.40% Mismatches: 1  
 Query Match: 98.81% Indels: 1  
 Db: 9 Gaps: 0

US-09-848-852a-3 (1-332) x HSM800198 (1-1782)  
 QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAlaHis-LeuAl 20  
 Db 320 ATGTACTCTGCCCAACGGTTTTGGGGGACAATCTGGCCAGAAAGGGAGCGACGCTTGC 379  
 QY 20 aProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40  
 Db 380 ACCCCCTGATGCGCAGCATCTCATCAGCAATGTGTGCACATCGGGACCATGTGGCCCA 439  
 QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGlu 60  
 Db 440 GGAGCTTTTTCAGGGCTCAGATTTGGGCATGTCAGAAAGAGGAGAGAGGCTGGGAGAA 499  
 QY 60 salAGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80  
 Db 500 AGCCGCCACGACAGCCCTCGAGAGGAGCATGTGACCTCGGTACAGAGCATCTTGA 559  
 QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGlu 100  
 Db 560 CGAATCTCTTAACAGTATGGCAGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGAA 619  
 QY 100 sLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGl 120  
 Db 620 GCTGGAGGACATTTTCCAGCAGGAGTTTTCCACCCCTTCCAGGAAGGGCCTGGTGTGCA 679  
 QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140  
 Db 680 GCTGATCCAGCTTTACCAGCGGATGCCAGCAATGCCATGGTGAGGGGCTCCGAGTGC 739  
 QY 140 aTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160  
 Db 740 TTATAGCGGACGCTGTGACCATGATGACCTGGGGACCTGTATGGACAGAACTGGCT 799  
 QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180  
 Db 800 CAATGACCAGGTGATCAACATGTATGGAGACCTGTGTGACACACAGTCCCTGAAAGGT 859  
 QY 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200  
 Db 860 GCATTTCTCAATAGTCTTCTATGATAAATCCGTACCAAGGGTTATGATGGGGTGAA 919  
 QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220  
 Db 920 AAGTGGACCAAAACGTGGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACCT 979  
 QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240  
 Db 980 GGAGGTGCATTTGGTCCCTCATCTCTGTGATGTAGGGCAGCAGCACCATCACCTATTTGA 1039  
 QY 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260  
 Db 1040 CTCGCGAGGTACCTTAAACCCCGCTGACCTAAGCATATATGCCAAGTATCTACAGGCAGA 1099  
 QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280  
 Db 1100 GCGGTGAAGAAGACCCGACTGGATTTCCACCAGGGCTGGAAGGTTACTTCAAAATGNA 1159  
 QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300  
 Db 1160 TGTGCCAGGCAGAAATAATGACAGTGACTGTGGTGTCTTTGTGTGCTACTGCAAGCA 1219  
 QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgGln 320  
 Db 1220 TCTGGCCCTGTCTCAGCCATTGACCTTACCACAGCAGGACATGCCCAAACCTTCGTGCGCA 1279  
 QY 320 nIleTyrLysGluLeuCysHisCysLysLysLeuThrVal 332  
 Db 1280 GATCTACAAGGAGCTGTGCTACTGCAAACTCACTGTG 1316  
 RESULT 2  
 AX081143



LOCUS AX081143 2206 bp DNA linear PAT 27-FEB-2001  
DEFINITION Sequence 9 from Patent WO0109292.  
ACCESSION AX081143  
VERSION AX081143.1 GI:13170039  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 2206)  
TITLE Gong, L. and Yeh, E.T.  
JOURNAL Composition and methods relating to senp1 - a sentrin-specific  
protease  
PATENT: WO 0109292-A 9 08-FEB-2001;  
Board of Regents, The University of Texas System (US)  
FEATURES  
Location/Qualifiers  
source  
1..2206  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 491 a 612 c 605 g 498 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.08e-164 Length: 2206  
Score: 1751.00 Matches: 331  
Percent Similarity: 99.40% Conservative: 0  
Best Local Similarity: 99.40% Mismatches: 1  
Query Match: 98.81% Indels: 1  
DB: 6 Gaps: 0  
US-09-848-852A-3 (1-332) \* AX081143 (1-2206)  
Qy 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20  
Db 759 ATGTACTGTCCCAACGGTTTGGGGACAATCTGGCGAAGAGGGGAGCGAGCTTGGC 818  
Qy 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG1 40  
Db 819 ACCCCTGATGCCAGCATCTCATCAGCAATGTGTGCAGCATCGGGACCATGTGGCCCA 878  
Qy 40 pGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60  
Db 879 GGAGCTTTTTCAGGCTCAGATTGGGCATGCGCAGAGGAGCAGAGAGCCCTGGGGAGAA 938  
Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80  
Db 939 ACCCGCCAGCAGACCCCTCGCAGAGAGCATGTGACCTGCGTACAGCATCTTGGG 998  
Qy 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValAlaGluLy 100  
Db 999 CGAATCTCTCAAGGTATGCGAGCCTCATACCCCTCAGCACTGATGAGGTAGTAGAGAA 1058  
Qy 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuG1 120  
Db 1059 GCTGAGAGACATTTTCCACAGAGAGATTTCACCCCTTCCAGGAAGGCGCTGGTGTGCA 1118  
Qy 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140  
Db 1119 GCTGATCCAGCTTCTACCAAGCGATGCCAGGCAATGCCATGTGAGGGCTTCCGAGTGGC 1178  
Qy 140 aTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160  
Db 1179 TTATAAGCGACGCGTGCACCATGGATGACTTGGGGACCTGTATGGACAGAACTGGCT 1238  
Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLySva 180  
Db 1239 CAATGACAGGTGATGAACATGTATGGAGACGTGCTCATGCACACAGTCCCTGAAAGGT 1298  
Qy 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLy 200  
Db 1299 GCATTTCTTCAATAGTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGGTAA 1358  
Qy 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220

|||||  
Db 1359 AAGGTGGACAAAACGTTGGACATCTTCAATAAGGAGCTACTGTCTAAATCCCATCCACT 1418  
Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240  
Db 1419 GGAGGTGATTTGGTCCCTCATCTCTGTGTGATGTGAGGCGACGACCATCATATTTGA 1478  
Qy 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaG1 260  
Db 1479 CTCGAGCGTACCTTAACCCGCTGCCTTAAGCATATTGCCAAGTATCTACAGCGAGA 1538  
Qy 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280  
Db 1539 GCGGTAAAGAAAGACCGACTGGATTTCACAGCGCTGGAAAGGCTTACTTCAAAATGAA 1598  
Qy 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300  
Db 1599 TGTGGCCAGGAGCAATAATGACAGTGTGGTGTCTTTGTGTGCTAGTACTGCAAGCA 1658  
Qy 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgG1 320  
Db 1659 TCTGGCCCTGTCTCAGCCATTTCAGCTTCAACGAGGACATGCCCAACTTCGTCGGCA 1718  
Qy 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332  
Db 1719 GATCTACAAGGAGCTGTGTCTACTGCAAACTCACTGTG 1755  
RESULT 3  
AY008763  
LOCUS  
DEFINITION Homo sapiens sentrin/SUMO-specific protease (SEN3), mRNA, complete  
cds.  
ACCESSION AY008763  
VERSION AY008763.1 GI:11245810  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 2224)  
AUTHORS Yeh, E.T., Gong, L. and Kamitani, T.  
TITLE Ubiquitin-like proteins: new wines in new bottles  
JOURNAL Gene 248 (1-2), 1-14 (2000)  
MEDLINE 20267842  
PUBMED 10806345  
REFERENCE 2 (bases 1 to 2224)  
AUTHORS Gong, L. and Yeh, E.T.  
TITLE Direct Submission  
JOURNAL Submitted (04-OCT-2000) Institute of Molecular Medicine, University  
of Texas-Houston, 2121 W. Holcombe Blvd. 928, Houston, TX 77030,  
USA  
FEATURES  
Location/Qualifiers  
source  
1..2224  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17p13"  
1..2224  
/gene="SEN3"  
52..1776  
/gene="SEN3"  
/codon\_start=1  
/product="sentrin/SUMO-specific protease"  
/protein\_id="AAG33252.1"  
/db\_xref="GI:11245811"  
/translation="MKETIQGTGSGWPPPPGIPPAYSSPRDGLRWPPPPKPKRLKS  
GGGPGDPGGTTPTRRLPAPRFASASEEEDDEDEDEDEEVEVAWRLPPRW  
SGLGTSORPALRPSHRKTCORERRAMRAFOLLYSKSTSLTFHKLWGRHRRRG  
LAHPKNHLSFQGGATQVPSPCRFSPGPPPRGLGLGALMAEDGVGRSPVPSG  
PPMEDGLRWTPKSPDLGSLJCTLPNGFGSGGSGEGERSLAPPDASILLINCS1  
FHWAEQELFGSDLGMAEAEAPCEKAGHSPLREHVTVCQSLDLEFLQYGLSLPL  
STDVEVKLEDIFQOEFSTPSRKGLVLQLIQSVQRMFGNAMVRGFRVAYKRHLTMD  
LGTLYGQNLNDQVMNMYGDLVMDTVPEKVHFFNSFFYDKLRTKGYDGVKRWTKNVDI

FNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLOAEAVKKDR  
LDFHQGWKGYFKMNVARQNDSDCGAFVLQYCKHLALSQPFSTQDDMPKLRRIYKE  
LCHCKLTV"

BASE COUNT 502 a 615 c 612 g 495 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,09e-164 Length: 2224  
Score: 1751.00 Matches: 331  
Percent Similarity: 99.40% Conservativeness: 0  
Best local Similarity: 99.40% Mismatches: 1  
Query Match: 98.81% Indels: 1  
DB: 9 Gaps: 0

US-09-848-852A-3 (1-332) x AY008763 (1-2224)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20  
Db 777 ATGTACTCTGCCACACGGTTTGGGGACAATCTGGCCAGAGGGGAGCGAGCTTGGC 836  
QY 20 aProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40  
Db 837 ACCCTGTATGCGAGCATCTCATCAGCATGTGTGCGCCAGAGGGGAGCGAGCTTGGCCCA 996  
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60  
Db 897 GGAGCTTTTTCAGGCTCAGATTTGGGCATGCGCAGAGAGGAGGCGCTGGGCGAGAA 956  
QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80  
Db 957 AGCCGCGCCAGCAGACCCCTCGGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGA 1016  
QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValGluLy 100  
Db 1017 CGAATTCCTCAACAGTATGCGAGCTCATACCCCTCAGACATGATGAGTAGTAGAGAA 1076  
QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGl 120  
Db 1077 GCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAGGCGCTGGTGTGCA 1136  
QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140  
Db 1137 GCTGATCCAGCTTACACGGATGCGCAGCAATGCCATGCTGGTGGAGGGGCTCCGAGTGGC 1196  
QY 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyClnAsnTrpLe 160  
Db 1197 TTATAAGCGGACGCTGCTGACCATGATGATGCTGGGAGCTTGTATGGACAGCACTGGCT 1256  
QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180  
Db 1257 CAATGACAGGTCATGAACATGATGACACCTGCTGATGGACACAGTCCCTGAAGAAGT 1316  
QY 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200  
Db 1317 GCATTTCTTCAATAGTTTCTTATGATAAATCCCGTACCAAGGGTTATGATGGGTGAA 1376  
QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuIleProIleHisLe 220  
Db 1377 AAGGTGGACAAAGAGTGGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACCT 1436  
QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240  
Db 1437 GGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTGAGGCGACGACCATCACCTATTTTCA 1496  
QY 240 pSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260  
Db 1497 CTGCGACGCTACCTTAACCGCGCTGCGCTAAGCATATTGCCAAGTATCTACAGCAGA 1556  
QY 260 uAlaValLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280  
Db 1557 GCGGGTAAAGAAGACCGACTGATTTCCACGAGGCTGGAAGGTTACTTCAAAATGAA 1616  
QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysH 300

Db 1617 TGTGGCCAGGCAATAATGACAGTGCACGTGTGTGCTTTTGTGTTGCTAGTACTGCAAGCA 1676  
QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgGl 320  
Db 1677 TCTGGCCCTGTCTCAGCCATTGAGCTTCACCCAGCAGGAGACATGCCCAAACTTCGTGCGCA 1736  
QY 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332  
Db 1737 GATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTG 1773  
RESULT 4  
AF199459  
LOCUS Homo sapiens SUMO-1 specific protease 3 (SSP3) mRNA linear PRI 31-OCT-2001  
DEFINITION Homo sapiens SUMO-1 specific protease 3 (SSP3) mRNA, complete cds.  
ACCESSION AF199459  
VERSION AF199459.1 GI:16550942  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2258)  
AUTHORS Choi,S.J., Jeon,Y.J., Kim,K.I., Nishimori,S., Suzuki,T., Uchida,S.,  
Shimbara,N., Tanaka,K. and Chung,C.H.  
TITLE Direct Submission  
JOURNAL Submitted (28-OCT-1999) Molecular Biology, Seoul National  
University, Shillim-dong san 56-1, Seoul 151-742, Korea  
FEATURES  
source  
1..2258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
gene  
1..2258  
/gene="SSP3"  
CDS  
71..1795  
/gene="SSP3"  
/codon\_start=1  
/product="SUMO-1 specific protease 3"  
/protein\_id="AAL25652.1"  
/db\_xref="GI:16550943"  
/translation="MKETIQGTGSGWPEPPGPIPPAYSSPRERLRWPPPPKPRKLS  
GGGPDGPGSTTPARLPVPRFSFASASEEEEEEEEEDEEEDEEEVAWRLPPRW  
SQLGTSQRPSRPTHRKTCQRRAFRMLYKSTSLTFHWKLGRRHRRRG  
LAHPKHLSPQGGATPOVPRCCRFSPRPPRLGLGALMAEDVRGSPVPSG  
PPMEDGLRWTPKSPLDPSGLISCTLPNGFGGSGPGRSLAPPDAS ILISNCSI  
GDHVAQELFGSDLGMAEERPEKAGQHSPLREHVTQVSLDEFLQYSGILPL  
STDEVEKEDIFQEEFTSPSRKGLVLQISYQRMGNAMVRFVAYKRHLVTMD  
LGTLYGQNLNDQVMNMYGDLVMDTPVPEKHFFNSFFYDKLTGTGDKGVRKWTNVDI  
FNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLOAEAVKKDR  
LDFHQGWKGYFKMNVARQNDSDCGAFVLQYCKHLALSQPFSTQDDMPKLRRIYKE  
LCHCKLTV"  
BASE COUNT 519 a 626 c 618 g 495 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,12e-164 Length: 2258  
Score: 1751.00 Matches: 331  
Percent Similarity: 99.40% Conservativeness: 0  
Best local Similarity: 99.40% Mismatches: 1  
Query Match: 98.81% Indels: 1  
DB: 9 Gaps: 0  
US-09-848-852A-3 (1-332) x AF199459 (1-2258)  
QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20  
Db 796 ATGTACTCTGCCACACGGTTTGGGGACAATCTGGCCAGAGGGGAGCGAGCTTGGC 855  
QY 20 aProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40  
Db 856 ACCCTGTATGCGAGCATCTCATCAGCAATGTGTGACATCGGAGGAGCTTGGCCCA 915  
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60

```
|||||
Db 916 GGAGCTTTTTCAGGCTCAGATTTGGGCATGGCAGAGAGCAGAGAGCCCTGGGGAGAA 975
Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
Db 976 AGCCGGCCAGCAGCAGCCCTTCCGAGAGGACATGTGACCTGCCGTACAGACATCTTGGGA 1035
Qy 80 pGluPheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluValValGluLy 100
Db 1036 CGAATTCCTTCAACAGTATGCGCCCTCATACCCCTCAGCACTGATGAGGTAGTAGAGAA 1095
Qy 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGl 120
Db 1096 GCTGGAGGACATTTTCCAGCAGAGAGTTTCCACCCCTTCCAGGAAGGGCCCTGGTGTGCA 1155
Qy 120 nLeuIleGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 1156 GCTGATCCAGCTTATACCAAGCGGATGCCAGCAATGCCATGATGGAGGGCTTCCGAGTGGC 1215
Qy 140 aTyrglyArgHisValLeuThrMetAspLeuGlyThrLeuTyrglyGlnAsnTrpLe 160
Db 1216 TTATAAGCGGCACGTGTCACCATGGATGACTTGGGACCTTGTATGGACAGAACTGGCT 1275
Qy 160 uAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThrValProGluLysVa 180
Db 1276 CAATGACCAAGGTGATGAACATGTATGGAGACCTGCTCATGGACACAGTCCCTGAAAGGT 1335
Qy 180 lHisPhePheAsnSerPhePheTyrglyAspLysLeuArgThrLysGlyTyrglyValLy 200
Db 1336 GCATTTCTTCAATAGTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGGTGA 1395
Qy 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220
Db 1396 AAGGTGACCAAAACGTGGACATCTCAATAAGAGCTACTGTAATCCCATCCACCT 1455
Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrglyPheAs 240
Db 1456 GGAGGTGCATTTGGTCCCTCACTCTGTTGATGTGAGGCGGACCACTACCTATTTTGA 1515
Qy 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrglyLeuGlnAlaGl 260
Db 1516 CTCGCGGTACCTTAAACCCCGCTGCCCTTAAAGCATATTGCCAAGTATCTACAGGCAGA 1575
Qy 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrglyPheLysMetAs 280
Db 1576 GCGGTAAAGAAAGACCGACCTGGATTTCCACCGGCTGGAAGAGTTACTTCAAAAGTAA 1635
Qy 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrglyLysHI 300
Db 1636 TGTGGCCAGGAGAATAATGACAGTACTGTGTGTCTTTGTGTGAGTACTGCAAGCA 1695
Qy 300 sLeuAlaIleuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgGgl 320
Db 1696 TCTGGCCCTGTCTCAGGCATTTACCTGCTTACCCAGCAGACATGCCCAAACTTCGTCGCGA 1755
Qy 320 nIleTyrglyGluLeuCysHisCysLysLeuThrVal 332
Db 1756 GATCTACAGGAGCTGTGTCTACTGCAAACTCAGTGTG 1792

RESULT 5
HSM805332
LOCUS      2269 bp      mRNA      linear      PRI 12-JUL-2002
DEFINITION Homo sapiens mRNA; cDNA DKF2p762A152 (from clone DKF2p762A152).
ACCESSION AL834294
VERSION   AL834294.1 GI:21739883
KEYWORDS   .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; HomInidae; Homo.
REFERENCE   1 (bases 1 to 2269)
AUTHORS    Koehrer K., Beyer A., Mewes H.W., Weil B. and Wiemann S.
TITLE      Direct Submission
```

JOURNAL  
COMMENT

Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
sequenced by BMFZ (Biomedical Research Center at the Charite,  
Berlin/Germany) within the cDNA sequencing consortium of the German  
Genome Project.

This clone (DKF2p762A152) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further  
information about the clone and the sequencing project is available  
at <http://mips.gsf.de/proj/cDNA/>.

FEATURES  
source

1..2269  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="17"  
/clone="DKF2p762A152"  
/tissue\_type="melanoma (MeWo cell line)"  
/clone\_lib="762 (synonym: hm12). Vector pSport1; host  
DH10B; sites NotI + SalI"  
/dev\_stage="adult"  
1..2269  
/gene="DKF2p762A152"  
78..1802  
/gene="DKF2p762A152"  
/note="partial similarity to sentrin/SUMO-specific  
protease (Homo sapiens)"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD38967.1"  
/db\_xref="GI:21739884"  
/translation="MKETIOGTGSGWPPGPPGIPPAYSSPRRLRWPPPPKPLKLS  
GGFGPDGCGTTPVARRLPVPRSPDASASEEEEEDEDEEEVAARLPPPM  
SQLTGSRPRSPRTHKTKQRRRARAFRLMLLYSKSLTFHMLKMGHRRRG  
LAHPKHLSPQGGATQVPSGCCFDPGPPPLRLGLLGLMAEDVGRGSPVPSG  
PMEEADGLRTPKSLDPSGLLSCITLPGNGSGSGEGESLAPDASLISNVCST  
GDHVAQELFQSGDLGMAEAEERGERAGHSPLREHVTCTQSLDFLQTYGSLIPL  
STDVYVEKLEIDIFQOEPSTPSRKGLVLQIOSYORMPGNAVGRFVAYKRHVLTMD  
LGYLQGNWLNDOMNMVGLVMDTVEKVFHFNSTFYDKLRTKGYDGVKRWTKNVD  
FNKELLILPIHLEWHWSLISVDVRRRTITYDSQRTLNRRCPKPHIAKYLQAEVKKDR  
LDPHQMGKGYFKMNVARNQNDSDCGAFVLOYCKHALUSQPSFTQDMPKLRIYKE  
LCHCKLTIV"

## gene

1..2269  
/gene="DKF2p762A152"  
78..1802

## CDS

1..2269  
/gene="DKF2p762A152"  
/note="partial similarity to sentrin/SUMO-specific  
protease (Homo sapiens)"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD38967.1"  
/db\_xref="GI:21739884"

## polyA\_signal

2228..2233  
/gene="DKF2p762A152"

## polyA\_site

2252  
/gene="DKF2p762A152"

BASE COUNT 522 a 628 c 621 g 498 t

## ORIGIN

Alignment Scores:  
Pred. No.: 1..13e-164 Length: 2269  
Score: 1751.00 Matches: 331  
Percent Similarity: 99.40% Conservative: 0  
Best Local Similarity: 99.40% Mismatches: 1  
Query Match: 98.81% Indels: 1  
DB: 9 Gaps: 0

US-09-848-852a-3 (1-332) x HSM805332 (1-2269)

Qy 1 MetTySerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAlaHis-LeuAl 20

Db 803 ATGTACTCTGCCCAACGCTTTTGGGACAACTCTGGGCCAAGAGGAGCCAGCTTGGC 862

Qy 20 aProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40

Db 863 ACCCCCTGATGCCAGCATCTTCATCAGCAATGTGTGCAGCATCGGGACCATGTGGCCA 922

Qy 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60

Db 923 GGAGCTTTTTCAGGCTCAGATTTGGGGCATGGCAGAGAGCAGAGCCCTGGGGAGAA 982

Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80

Db 983 AGCGGCCACACAGCCCCCTGCGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGA 1042

QY 80 pGluPheLeuGlnThrTyrglySerLeuileProLeuSerThrAspGluValValGluLy 100

Db 1043 CGAATTCCTTCAAAAGTATGCGAGCTCATACCCCTCAGCAGCTGATGAGTAGAGAA 1102

QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuG1 120

Db 1103 GCTGGAGGACATTTCCAGCAGGAGTTTCCACCCCTCCAGGAGGCGTGGTGTGCA 1162

QY 120 nLeuileGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140

Db 1163 GCTGATCCAGCTTACAGCGGATGCGAGCAATGCCATCGTGGAGGGCTCCGAGTGGC 1222

QY 140 aTyrlYsArgHisValLeuThrMetAspLeuGlyThrLeuTyrglyGlnAsnTrpLe 160

Db 1223 TTATAAGCGCAGCTGCTGACCATGATGACTTGGGAGCTTGTATGGACAGAACTGGCT 1282

QY 160 uAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThrValProGluLysVa 180

Db 1283 CAATGACCAAGTATGAACATGTATGAGACCTGGTCATGGACACAGTCCCTGAAAGGT 1342

QY 180 lHisPheAsnSerPhePheTyrglyAspLeuArgThrLysGlyTyrglyVally 200

Db 1343 GCATTTCTTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGTTATGATGGGTGAA 1402

QY 200 sArgTyrThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220

Db 1403 AAGTGGACCAAAAGCTGGACATCTTCAATAGGAGCTACTGCTAATCCCATCCACCT 1462

QY 220 uGluValHisTrpSerLeuileSerValAspValArgArgThrIleThrTyrglyPheAs 240

Db 1463 GGAGGTGCATTTGGTCCCTCATCTCTGATGTGAGCGCAGCACCATCACCTATTTTGA 1522

QY 240 pSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrglyLeuGlnAlaG1 260

Db 1523 CTGCGACGTACCTAAACCGCGCTGCGCTAAGCATATTTGCCAAGTATCTACAGCAGA 1582

QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrglyPheLysMetAs 280

Db 1583 GCGCGTAAAGAAGACCGACTGATTTCCACGAGGCTGGAAGGTTACTTCAAAATGAA 1642

QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrglyCysLysHi 300

Db 1643 TGTGGCCAGCAGAAATATGACAGTACTGTGGTCTTTGTGTCAGTACTGCAAGCA 1702

QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgG1 320

Db 1703 TCTGGCCCTGTCTCAGCCATTGAGCTTCCAGCAGGACATGCCCAAACTTCGTGGCA 1762

QY 320 nIleTyrlYsGluLeuCysHisCysLysLeuThrVal 332

Db 1763 GATCTAAGGAGCTGTGCTGACTGCAAACTACTGTG 1799

RESULT 6

AF194031

LOCUS AF194031 1707 bp mRNA linear ROD 01-NOV-2000

DEFINITION Mus musculus SMT3 isopeptidase 1 (Smt3ip) mRNA, complete cds.

ACCESSION AF194031

VERSION AF194031.1 GI:11066007

KEYWORDS

SOURCE Mus musculus.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1707)

AUTHORS Nishida,T., Tanaka,H. and Yasuda,H.

TITLE A novel mammalian Smt3-specific isopeptidase 1 (SMT3IP1) localized in the nucleolus at interphase

JOURNAL Eur. J. Biochem. 267 (21), 6423-6427 (2000)

MEDLINE 20485517

PUBMED 11029585

REFERENCE 2 (bases 1 to 1707)

AUTHORS Nishida,T., Tanaka,H. and Yasuda,H.

TITLE Direct Submission

JOURNAL Submitted (13-Oct-1999) School of Life Science, Tokyo University of Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan

FEATURES

Location/Qualifiers

source

1..1707

/organism="Mus musculus"

/db\_xref="taxon:10090"

gene

1..1707

/gene="Smt3ip"

CDS

1..1707

/gene="Smt3ip"

/codon\_start=1

/product="SMT3 isopeptidase 1"

/protein\_id="AAG28418.1"

/db\_xref="GI:11066008"

/translation="MKETIQGTSWGPPEPGPTTYSNPRERLRWRLPKPRKSGG

FGFPPGSGTTPVTRRLPRPSPFDASASEEEEEDEEEVAWRLLPRLRWQLGAS

QSRALRPSHRKTCSSRRRRARAFQMLLYSKSTLTFHWKLMGRHRRRLAHPKN

HLSPQEGATPQVSPCCRFDSPRGLPPRLGLLALMAEDMGSPVPVSGPPMPEED

GLRWTPKSPDPDGLSLLCTLPNGFGGLSGPEGRSLAPPDASILLINVCISGDHVAQ

ELFSSDLGIAEADRTGEKAGQSPLEEHVTCVQSILDEFQTYIGSLIPLSTDEVV

EKLEDIFQOEFSTPSRKSLLQLTIQSYQRMFCNAMRGFRVSKRHVLTMDLSTLYG

QNWLDQVMNMYGLVMDTVPEKHFNSFYDKLRTKGYDGVKRWTKNVDIFNKELL

LIPHLVHWSLISVDVRRRTITVFDSQRTLNRRCPKHIAKYLOAEAVKKRDLDFHOG

WKGYFKMNVARQNDSDCGAFVLQYCKHLALSLSPFSFTQQDMPKLRROIYRELCHKL

TV"

BASE COUNT 406 a 477 c 461 g 363 t

ORIGIN

Alignment Scores:

Pred. No.: 2,28e-159 Length: 1707

Score: 1696.00 Matches: 320

Percent Similarity: 97.60% Conservative: 5

Best Local Similarity: 96.10% Mismatches: 7

Query Match: 95.71% Indels: 1

DB: 10 Gaps: 0

US-09-848-852A-3 (1-332) x AF194031 (1-1707)

QY 1 MetTyrlSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20

Db 708 ATGTACTCTACCAATGGTTTGGGGGACTGCTGCTGCTGAAGGGGAGCGAGCTGCG 767

QY 20 aProProAspAlaSerIleLeuileSerAsnValCysSerIleGlyAspHisValAlaG1 40

Db 768 ACCCCCTGATGCGCAGCATACTCATCAGTAATGTGTGCAGCATTTGGAGACCACGCTCA 827

QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60

Db 828 GGAATCTTTTCAGACTCTGACTTGGGCATTCACAAGAGGAGCATCGGACTGGGAGAA 887

QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80

Db 888 ACCTGGCCAGCATAGCCCTTCGGGAGGAGAACATGTGACCTGTGTGCAGAGTATCTAGA 947

QY 80 pGluPheLeuGlnThrTyrglySerIleLeuileProLeuSerThrAspGluValValGluLy 100

Db 948 TGAATTCCTTCAAACTTATGGCAGGCTCATCCCTCTCAGCAGCTGATGAAGTTGTAGAA 1007

QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuG1 120

Db 1008 GTTGGAGACATTTTCCAGCAGAGTCTCTACACCTCAAGAGAGAGTCTGGTACTACA 1067

QY 120 nLeuileGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140

Db 1068 GCTGATCCAGTCTTATCAGCGGATGCCAGGCAACGCTATGTTAAGGGGCTTCCGGGTATC 1127

QY 140 aTyrlYsArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrglyGlnAsnTrpLe 160

Db 1128 CTATAAGCGACACGTGCTCACCATGATGGTACCTTATATGGACAGAACTGGCT 1187

```

Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
Db 1188 CAATGACAGGTGATGAACATGATGAGATCTGCTGATGACACAGTCCCTGAAAGGT 1247
Qy 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
Db 1248 GCATTTCTTCAACAGCTTCTTCTATGATGATAAATCCGTACCAAGGGTTATGATGGGGTAAA 1307
Qy 200 sArgTTrpThrLysAsnValAspIlePheAsnLysLeuLeuLeuLeuLeuLeuLeuLeu 220
Db 1308 GAGGTGGACCAAAAATGTGGACATCTTCAATGAAGAAATCTGCTAATCCCATCCATCT 1367
Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240
Db 1368 GGAGGTGCACCTGGTCCCTTATCTAGTTGATGTAAGGCGACGTACCATCACCTATTITGA 1427
Qy 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaG 260
Db 1428 CTCCAGCGAACTCTAAATCCCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAGA 1487
Qy 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280
Db 1488 GGCAGTCAAAAAGACCGACGTGGACTCCATCAGGGCTGGAAAGGTACTTCAAAATGAA 1547
Qy 280 nValalaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysH 300
Db 1548 TGTGCCAGGCAGAAATATGACAGTACTGTGGTGGCTTTGTATTACAGTACTGCAAGCA 1607
Qy 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgG 320
Db 1608 CCTGGCCCTGTCTCAGCCATTCAGCTTCCAGCAGCAGACATGCCAAACTCCGTCGTCA 1667
Qy 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
Db 1668 GATCTACAGGAAGTGTGTCTACTGCAAACTCAGCTGTG 1704

RESULT 7
AY008764 2174 bp mRNA linear ROD 21-NOV-2000
LOCUS Mus musculus sentrin/SUMO-specific protease (SENp3) mRNA, complete
DEFINITION
ACCESSION AY008764
VERSION AY008764.1 GI:11245812
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL Yeh,E.T., Gong,L. and Kamitani,T.
MEDLINE Ubiquitin-like proteins: new wines in new bottles
20267842
Gene 248 (1-2), 1-14 (2000)
PUBMED
10806345
REFERENCE
AUTHORS Gong,L. and Yeh,E.T.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2000) Institute of Molecular Medicine, University
of Texas-Houston, 2121 W. Holcombe Blvd. 928, Houston, TX 77030,
USA
FEATURES
source
1. .2174
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/chromosome="11"
1. .2174
/gene="SENp3"
52. .1758
/gene="SENp3"
/codon_start=1
/product="sentrin/SUMO-specific protease"
/protein_id="AAG33253.1"

```

```

/db_xref="GI:11245813"
/translation="MKETIOGTSGWGPPEPGGTTYSNPRRERLRWLPKPKRKSQS
GFQDPGSGTIVTRRLPAPRPSPFADASAEDEEEDDEEVAARLPPRWGLGAS
QRSRALRPSHRKTCQRRRARAFQMLLYSKSLTFHWKLMGRHGRNRKRLAHPKN
HLSPEGGATQVPSPCCRFDSPRLGPPRLGLGALMAEDMGSPSPVPSGPPMEED
GLRWTPKSPLDPSGLSCTLPNGFGLSGPEGRSLAPDASLISNVCSIGDRVAQ
ELFOSDLGIAEADRTGERAGQSPLEREHVTCVQSITLDEFLOTYGSLLIPLSTDEVV
EKLEDI FQOEFTSPSRKSLVQLQILOSQYORMPGNMAVGRFVSYKRHVLTMDDICTLYG
ONWLNDOVMNYGDLVMDVTPEKVFHFFNSFYDKLRTKGYDGVKRWTKNVDIFNKELL
LPIHLHVHWSLISVDVRRRTIIFYDSQRLNRRCPKHIAYKLOAEAVKRDRLDFHQG
WKGYFKNMVARQNNDSDCGAFVIOGQKHLALSQPFPSFTQDMPKLRQIYKELCHCKL
TV"
BASE COUNT 513 a 585 c 570 g 506 t
ORIGIN
Alignment Scores:
Pred. No.: 3.22e-159 Length: 2174
Score: 1696.00 Matches: 320
Percent Similarity: 97.60% Conservative: 5
Best Local Similarity: 96.10% Mismatches: 7
Query Match: 95.71% Indels: 1
DB: 10 Gaps: 0
US-09-848-852A-3 (1-332) x AY008764 (1-2174)
Qy 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAlaHis-LeuAl 20
Db 759 ATGTACTCTACCAATGGTTTGGGGACTGTGTGGTCTCTGAAGGGAGCGCAGCTGGC 818
Qy 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG 40
Db 819 ACCCCTGATGCCAGCATCTCATCAGTAATGTGTGACGATTTGGAGACCAGCTGGCTCA 878
Qy 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60
Db 879 GGAACCTTTTCAGAGCTCTGACTTGGGCATTGCGAAGAGGAGCATGCGAGTGGGAGAA 938
Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
Db 939 AGCTGGCCAGCATAGCCCTTCGGGAGAGAACATGTGACCTGTGTGACAGATATCTTAGA 998
Qy 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValAlaGluLy 100
Db 999 TGAATTCTTCAAACTTATGCGAGCTCATCCCTCTCAGCAGTGTATGAAGTTGTAGAGAA 1058
Qy 100 sLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuG 120
Db 1059 GTTGGAGACATTTTCCAGCAGGAGTTCTCTACACCCCTCAAGGAGAGTCTGGTACTACA 1118
Qy 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 1119 GCTGATCCAGCTTATATCAGCGGATGCCAGGCAACGCTATGTTAGGGGCTTCCGGGTATC 1178
Qy 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160
Db 1179 CTATAAGCGACACGCTGCTCACCATGATGACTTGGGTACCTTATATGACAGAACTGGCT 1238
Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
Db 1239 CAATGACAGGTGATGAACATGTATGGAGATCTGTGATGGACACAGTCCCTCGAAAGGT 1298
Qy 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
Db 1299 GCATTTCTTCAACAGCTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGGTAAA 1358
Qy 200 sArgTTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuLeuLeuLeuLeu 220
Db 1359 GAGGTGGACCAAAAATGTGGACATCTTCAATAAGGAATTTACTGTCTAATCCCATCCATCT 1418
Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240
Db 1419 GGAGGTGCACCTGGTCCCTTATCTCAGTTGATGTAGGCGAGGTACCATCACCTATTTTGA 1478

```

```

QY 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaG1 260
|||||
Db 1479 CTCCAGCGAACTCTAAATCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAGA 1538
|||||
QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTyrLysGlyTyrPheLysMetas 280
|||||
Db 1539 GGCAAGTCAAAAAAGACCGACTGACATCCATCAGGCTGGAAAGGTTACTTCAAAATGAA 1598
|||||
QY 280 nAlaAlaArgGlnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300
|||||
Db 1599 TGTGGCCAGCAGCAATAATGACAGTACACTGTGTGCTTTGTATTACAGTACTGCAGCA 1658
|||||
QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgG1 320
|||||
Db 1659 CCTGGCCCTCTCTCAGCAATTCAGCTTACCCAGCAGGACATGCCCAAACTCGTGTCA 1718
|||||
QY 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
|||||
Db 1719 GATCTACAGGAAGTGTGTCTACTGCAAACTCACTGTG 1755
|||||
RESULT 8
AC126239 139405 bp DNA linear HTG 04-JUL-2002
LOCUS
DEFINITION
Felis catus clone RP86-474H17, WORKING DRAFT SEQUENCE, 12 unordered
pieces.
ACCESSION
AC126239.1 GI:21693903
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
cat.
SOURCE
Felis catus
ORGANISM
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
REFERENCE
1 (bases 1 to 139405)
AUTHORS
Benjamin,B., Blakesley,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Latic,P.,
Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Marquies,E.H., Mastello,C., Maskeri,B., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Paguirican,C., Pearson,R.,
Portnov,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 139405)
Green,E.D.
Direct Submission
Submitted (04-JUL-2002) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: daa
Center clone name: 474H17
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 131028 bases at least Q40
Consensus quality: 132577 bases at least Q30
Consensus quality: 133578 bases at least Q20
Insert size: 126000; agarose-fp
Insert size: 138305; sum-of-contigs
Quality coverage: 7.22x in Q20 bases; agarose-fp
Quality coverage: 6.57x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2527: contig of 2527 bp in length
2528 2627: gap of unknown length
* 2628 5297: contig of 2670 bp in length
* 5298 5397: gap of unknown length
* 5398 7601: contig of 2204 bp in length
* 7602 7701: gap of unknown length
* 7702 9908: contig of 2207 bp in length
* 9909 10008: gap of unknown length
* 10009 12523: contig of 2515 bp in length
* 12524 12624: gap of unknown length
* 12624 16221: contig of 3598 bp in length
* 16222 16321: gap of unknown length
* 16322 24505: contig of 8184 bp in length
* 24506 24605: gap of unknown length
* 24606 35941: contig of 11336 bp in length
* 35942 36041: gap of unknown length
* 36042 47449: contig of 11408 bp in length
* 47450 47549: gap of unknown length
* 47550 66243: contig of 18694 bp in length
* 66244 101268: contig of 34925 bp in length
* 101269 101368: gap of unknown length
* 101369 139405: contig of 38037 bp in length.
FEATURES
Location/Qualifiers
1..139405
/organism="Felis catus"
/db_xref="taxon:9685"
/clone="RP86-474H17"
/clone_lib="RP86"
1..2527
/note="assembly_fragment"
2628..5297
/note="assembly_fragment"
5398..7601
/note="assembly_fragment"
clone_end:SP6
vector_side:left
7702..9908
/note="assembly_fragment"
10009..12523
/note="assembly_fragment"
12624..16221
/note="assembly_fragment"
16322..24505
/note="assembly_fragment"
24606..35941
/note="assembly_fragment"
36042..47449
/note="assembly_fragment"
47550..66243
/note="assembly_fragment"
66344..101268
/note="assembly_fragment"
101369..139405
/note="assembly_fragment"
BASE COUNT 32325 a 36788 c 37913 g 31272 t 1107 others
ORIGIN
Alignment Scores:
Pred. No.: 3.07e-84 Length: 139405
Score: 969.50 Matches: 320
Percent Similarity: 22.51% Conservative: 3
Best Local Similarity: 22.30% Mismatches: 9
Query Match: 54.71% Indels: 1107
DB: 2 Gaps: 8
US-09-848-852A-3 (1-332) x AC126239 (1-139405)

```









Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysVal----- 75  
|||||  
Db 87138 GGCTGGCCAGCACAGCCCTCGCGGAGGAGCATGTGACCTGCGTACAGAGTAAGGGACC 87079  
75 ----- 75  
Db 87078 CTAGAGACCCAGTCTCTTCTTAGTAGCCAACTCTTAAGTTGGAATCTGGAGCCCTCAC 87019  
Qy 76 -----GlnSerIleLeuAspGluPheLeuGlnThrTyrGlyse 88  
||| |||||  
Db 87018 GCCACTCTCTTCCCTCCCTCAGGC-ATCTTGGATGAATTCCTCAAACTATTGGCAG 86960  
Qy 88 rLeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGlnI 108  
|||||  
Db 86959 CCTCATCCCTCTTAGCACTGATGAGTGGTAGAGAAATTGGAGGACATTTTTTCAGCAGGA 86900  
Qy 108 uPheSerThrProSer----- 113  
|||||  
Db 86899 GTTCTCTACACCTTCCAGGTGAGGCATGAAGAGAGATGTCGTAGAGAGGGCTGGGCCCT 86840  
Qy 113 ----- 113  
Db 86839 GAGAGTATGAGAGGTTGCTGCTGCCCTTCTCCCATAGGGATGTACAAGGCGGATGGAG 86780  
Qy 113 ----- 113  
Db 86779 GAAGGTAGAACTGAGGGGAGGAACCTGTAGGCAGAGTGCCACGCCCTGTAAACTGACT 86720  
Qy 113 ----- 113  
Db 86719 GGGGTGGAGTCTTTACCTGAGATCCTGCAATATTATGATATACCTGATACTCTGTGCCCT 86660  
Qy 113 ----- 113  
Db 86659 TCTGGAGACTAGGCCCTTTACAGGTCCTCAAAAGGATGTGTGCCCAAGAAAGGATTAG 86600  
Qy 113 ----- 113  
Db 86599 TTTTGTGTTGAAGTGGTTCCACACCCAGTGTGAAGCCCAACACAGGGCTTGAACCTACG 86540  
Qy 113 ----- 113  
Db 86539 ACCTTGAGATCAATACCTGAGCTGAGATCAGGAATCGGATGTTTAAACGACTGGGCCACC 86480  
Qy 113 ----- 113  
Db 86479 TAGGTGCCCCCACCAGGTGCCCTTATCTTAAGAATAAGAACTAGTTTAGGTCTT 86420  
Qy 113 ----- 113  
Db 86419 ATTCTGGCAGTAGAAAAGTACTTCTGTATGCTCTAGCCCCCTTGTCTCTGTGTGACTC 86360  
Qy 114 -----ArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMe 128  
|||||  
Db 86359 CACCTTGGCCTACTCAGGAAGGCCGTGTGCTGACGTGATCCAGTCTGATCCAGTCCAGCGGAT 86300  
Qy 128 tProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThrMe 148  
|||||  
Db 86299 GCCAGCAATGCCATGGTGAGGGGTTTCCGAGTGGCCTATAAGCGGCATGTACTGACCAT 86240  
Qy 148 tAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAsp----- 162  
|||||  
Db 86239 GGATGACCTGGGACCTTGTATGGACAGAACTGGCTCAATGACCAGGTGAGGAGTAGTGG 86180  
Qy 162 ----- 162  
Db 86179 AGAAATGGGCCAAAAGGGGATTTAGGAGAGCGGTCTCTGGGCCCTCTGTTTGGGGGA 86120  
Qy 163 -----GlnValMe 165  
|||||  
Db 86119 GCCCTATACCTATGCTGCACTCTCCATGGCAAGTGGCTCCCAATTTCTCCCCAGGTGAT 86060  
Qy 165 tAsnMetTyrGlyAspLeuValMetAspThrValProGluLys----- 179

Db 86059 GAACATGTATGGAGACCTGGTTCATGTGACAGCTCCCTGAAAGGTAGGCCCAACACGAGT 86000  
Qy 179 ----- 179  
Db 85999 ACCTCAGTCCCCAGAAAGCCCTTCTGCAGTTTGAAGCAGCTTTTTCAGTCCCCCTTTATCTC 85940  
Qy 179 ----- 179  
Db 85939 TTCATCCATAAAGAGAGATCTTGGTTTGGGGGAGAGGTGGTAGAAGGCAGTCACTTAAT 85880  
Qy 179 ----- 179  
Db 85879 CTGTAATCTTCTCTTTTAAAGTTTCTTATTTGAGAGAGAAAGGAGGGGGGAGG 85820  
Qy 179 ----- 179  
Db 85819 GAGGGAGAGTGGGGGGAAGGCAGGAAATCTCCAGAGATTCCCTGCTGAGTGCAGAGCC 85760  
Qy 179 ----- 179  
Db 85759 CAACGAGGGGCTTGAGCTCTCAATACCGAGGGCTTGAGCTCTCAATACCGAGATCACAA 85700  
Qy 179 ----- 179  
Db 85699 CCCACGCCGAACCAAGAGTCAAGACTCAACTGAGTCACTGAGTCACTGAGTCACTCAAA 85640  
Qy 179 ----- 179  
Db 85639 TTTGTAATCGTGCTGCTGAACTTACTTAATCTTTTCCCCAGTGTAGGGTTTTTCCCTGT 85580  
Qy 179 ----- 179  
Db 85579 GGTAACTCCACAATTTGGATGGGAAGCAGTAAGATAAGTTAAAGAGCTTTGTTGTTT 85520  
Qy 179 ----- 179  
Db 85519 TTACCCTCAGTGTGAGTGCAGGTTATTGAAACTGGCCCTTGGAGCCCTTAATGAAAG 85460  
Qy 179 ----- 179  
Db 85459 GAAGCAAGCGCCAGGTGGAGGGGACCTGGCAAGGGGCCCTGAGCTCTTTCAGTTCCAT 85400  
Qy 180 -----ValHisPhePheAsnSerPheTyrAspLysL 191  
|||||  
Db 85399 CCAGGTCTTTTGTATCATTTGGCACAGGTGCATTTCTTCAACAGTTTCTTCTATGATAAAC 85340  
Qy 191 euArgThrLysGlyTyrAspGlyValLysArgTrpThrLysAsn----- 205  
|||||  
Db 85339 TCCGCACCAAGGTTATGATGGGTGAAAGGTGGACCAAAACGATGATTAATCAATTCA 85280  
Qy 205 ----- 205  
Db 85279 TATACGCTGGTGTGACACCTTGCTCTGAGAGGATTAGAGTTCTGTCTTAGSAGCCCTT 85220  
Qy 205 ----- 205  
Db 85219 CCTAAAGCCCTGCTCACTCAGTCTTTCAAGTATTATAAGGCTCTGGTATGTGTGAGGTA 85160  
Qy 205 ----- 205  
Db 85159 GTCGTAGGCACAAGGATTTCAGTGTGTGAGTAAGGAAACCTTGCCCTTTATGGAGTTAAGG 85100  
Qy 205 ----- 205  
Db 85099 TAAATAAATAGTTAAATTTTCAGTGTGTAGTGTATGAATCACTCACTGAAAGGTAAT 85040  
Qy 205 ----- 205  
Db 85039 GCAGTAGCCCCGGGAGCTAGGCCCTTTAGATTGGGGGTTTGGGAGGCCCTCTCTCTCTG 84980  
Qy 205 ----- 205

Qy	205	-----	205
Db	83839	CATCCTGGGGTCTCAGGTTGTGTGTGACAGTAGGTTCTTGACCTGCCTTCTTCTTCCTCAC	83780
Qy	205	-----	205
Db	83779	CCCTGGAAATCACACATGCTGAACTTCACCTCAGTGGAGGAGCTCAGCTGTGTGCTTG	83720
Qy	205	-----	205
Db	83719	CCCCTGGGCTCCTTGTCTCGCCTCCACATCTTTAGGACCCCTCGCGGCAATGGCAGGTG	83660
Qy	206	-----ValAspTLeuPheAsnLys	211
Db	83659	GCTCCTGCTGTCCAACTGACCTCTGACACACACCTCCTTGGTGGGACATCTTCAATAG	83600
Qy	212	GluLeuLeuLeuIleProIleHisLeuGluValHisTrrSerLeuIleSerValAspVal	231
Db	83599	GAGCTCCTGCTAAATCCCATCCACCTGAGGTGCACTGGTCCCTCATCTCTGTTCACGTG	83540
Qy	232	ArgArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgCysProLys	251
Db	83539	AGGGACACCATCATCCTACTTTGACTCGCAGCGCACCTAAACCGCGCTGCCCTAA	83481
Qy	251	-----	251
Db	83480	GTTTGACAGGAGCGGGGGGGTGGGGGGGGTGGAGAAATCCAGTGGCAGGCGAT	83421
Qy	251	-----	251
Db	83420	AGTGGGCAAGAAGGCGAGGCTGGTGCCTGGGACCACCTATACCTAGGTTCACTTGGGAAA	83361
Qy	252	-----HisIleAlaLysTyrLeuGlnAla	259
Db	83360	CCGAGGCATCACTTCCTCTTTTCTGCTCCATAGCATATTGCCAAGTACCTACAGGCA	83301
Qy	260	GluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrrLysGlyTyrPheLysMet	279
Db	83300	GAAGCAGTGAGAAAGACCGGCTGGATTTCCACCAGGGCTGAAAGGTTACTTCAAATG	83241
Qy	279	-----	279
Db	83240	GTAAGTGTCTGGAGGAGGGGTACAGTGGTGATGGTGGTAGAGAGCGGTAGGAGG	83181
Qy	279	-----	279
Db	83180	AAGAAGTGAAGCCACACCCCGAGGAGCGTCCAGGTGAAGGACTCTCTTTTATTCTCTA	83121
Qy	280	-AsnValAlaArgGlnAsnAsnSerAspCysGlyAlaPheValLeu	295
Db	83120	GAAGCTGGCGGACAGAAATAATGACAGTGACTGTGGTGCTTGTGTACAGGTAAAGCAG	83061
Qy	295	-----	295
Db	83060	AGATCGGTGGGTTAAAGGTTTGGTAAGGAGGTGGGAGCGTGTCCACCTCTCTGGTTTG	83001
Qy	295	-----	295
Db	83000	GCTCACAGGCACCTTGTGAAGCAGAGAGCAACCTTAGGTGTTAGTGTGCAGAAACTCGGAT	82941
Qy	295	-----	295
Db	82940	CCCCTTGTCTCCCGGCTTCCCGGGCCCCACACTTGGCCTGGGTCTCTTGGTCTCTGGTC	82881
Qy	295	-----	295
Db	82880	TCTGCCCTGGTACAGTTGCTTCCCAACCCCGAGTGCAAGCATTTTCTTCAGGTATCGA	82821
Qy	295	-----	295
Db	82820	GAACAGTGACTCTAGAGTCAGGCTTTAGGCCTTAGGCCTATGGGAGCTTGGCTTCTGTGT	82761

```

Qy 295 ----- 295
Db 82760 GAAATGTGGATGTTCTGTGAAAGGTTTGGTGAGGTGGATCGCTCTGTGCAGCACAGC 82701
Qy 295 ----- 295
Db 82700 ACAGGTCTGACCTGGGGGATGCTCAGTAGATGGTTCTTATATGCGCTGCTTTGTTTAACAC 82641
Qy 296 -GlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetPr 315
Db 82640 CCAGTACTGACGACCTGGCCCTGTCTCAGCCATTCAGTTCACCCAGGACGACATGCC 82581
Qy 315 oLysLeuArgGlnIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
Db 82580 CAAACTTCGTGGCAGATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTG 82529

RESULT 10
AC126921/c
LOCUS
DEFINITION
  Bos taurus clone RP42-45D24, WORKING DRAFT SEQUENCE, 12 unordered
  pieces.
ACCESSION
  AC126921
VERSION
  AC126921.1 GI:21724098
KEYWORDS
  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
  cow.
ORGANISM
  Bos taurus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovidae; Bovinae; Bos.
  1 (bases 1 to 153553)
  BenJamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
  Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
  Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
  Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
  Margulies,E.H., Masello,C., Maskeri,B., Mastrian,S.D.,
  McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
  Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
  Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
  Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  NISC Comparative Sequencing Initiative
  Unpublished
  2 (bases 1 to 153553)
  Green,E.D.
  Direct Submission
  Submitted (10-JUL-2002) NIH Intramural Sequencing Center, 8717
  Grovemont Circle, Gaithersburg, MD 20877, USA
  ----- Genome Center
  Center: NIH Intramural Sequencing Center
  Center code: NISC
  Web site: http://www.nisc.nih.gov
  Contact: nisc_zoo@hgri.nih.gov
  ----- Project Information
  Center project name: ddi
  Center clone name: 045D24
  ----- Summary Statistics
  Sequencing vector: plasmid; n/a; 100% of reads
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Assembly program: Phrap; version 0.990319
  Consensus quality: 146066 bases at least Q40
  Consensus quality: 147748 bases at least Q30
  Consensus quality: 148824 bases at least Q20
  Insert size: 151000; agarose-fp
  Insert size: 152453; sum-of-contigs
  Quality coverage: 8.80x in Q20 bases; agarose-fp
  Quality coverage: 8.72x in Q20 bases; sum-of-contigs
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 12 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence

```

```

* as soon as it is available and the accession number will
* be preserved.
* 1
* 2312: contig of 2312 bp in length
* 2313: gap of unknown length
* 2412: contig of 3429 bp in length
* 2413: gap of unknown length
* 5841: gap of 3429 bp in length
* 5842: gap of unknown length
* 5942: contig of 2494 bp in length
* 8435: gap of unknown length
* 8436: gap of unknown length
* 8535: contig of 7264 bp in length
* 15799: gap of unknown length
* 15800: gap of unknown length
* 15899: gap of unknown length
* 25224: contig of 9325 bp in length
* 25225: gap of unknown length
* 25324: gap of unknown length
* 25325: contig of 7180 bp in length
* 32604: gap of unknown length
* 32605: contig of 8366 bp in length
* 40970: gap of unknown length
* 40971: gap of unknown length
* 41071: gap of unknown length
* 56590: contig of 15520 bp in length
* 56591: gap of unknown length
* 56592: gap of unknown length
* 56593: contig of 17079 bp in length
* 73769: gap of unknown length
* 73770: gap of unknown length
* 73870: contig of 16990 bp in length
* 90859: gap of unknown length
* 90860: gap of unknown length
* 111428: contig of 20469 bp in length
* 111429: gap of unknown length
* 111529: contig of 42025 bp in length.

FEATURES
  Location/Qualifiers
    1..153553
      /organism="Bos taurus"
      /db_xref="taxon:9913"
      /clone="RP42-45D24"
      /clone_lib="RP42"
    1..2312
      /note="assembly_fragment"
    2413..5841
      /note="assembly_fragment"
    vector_end:T7
    vector_side:left
    5942..8435
      /note="assembly_fragment"
    8536..15799
      /note="assembly_fragment"
    15900..25224
      /note="assembly_fragment"
    25325..32504
      /note="assembly_fragment"
    32605..40970
      /note="assembly_fragment"
    41071..56590
      /note="assembly_fragment"
    56691..73769
      /note="assembly_fragment"
    73870..90859
      /note="assembly_fragment"
    90960..111428
      /note="assembly_fragment"
    111529..153553
      /note="assembly_fragment"
    clone_end:SP6
    vector_side:right
  BASE COUNT 39178 a 37393 c 36616 g 39259 t 1107 others
  ORIGIN

Alignment Scores:
  Pred. No.: 3,23e-72 Length: 153553
  Score: 849.50 Matches: 285
  Percent Similarity: 22.21% Conservative: 4
  Best Local Similarity: 21.91% Mismatches: 11
  Query Match: 47.94% Indels: 1001
  DB: 2 Gaps: 7

US-09-848-852a-3 (1-332) x AC126921 (1-153553)
Qy 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAla-HisLeuAl 20

```



QY 205 ----- 205  
Db 34312 TTAACCTCCTGTTTCGGCGATGGAAGGGGCAAGAGTGAAGCAGGACAGACAAACAGA 34253  
QY 205 ----- 205  
Db 34252 GTAGCTGAGAAGTCCTGATGAGAAACTTGGACCAGAGTGGTCAGTGTGGAATATTGATA 34193  
QY 205 ----- 205  
Db 34192 GTAGTAGTTTTATTATTAGGTAGCTCTTGAGGTAGAACCAATAGGACTTTCTGATGGATT 34133  
QY 205 ----- 205  
Db 34132 GGACATGAGAGATGAGGAAAAAGATAGCAATCAAGGATGATGCTAGGTTTTTGGCCCTAAG 34073  
QY 205 ----- 205  
Db 34072 TTACTGGGAATATAATTAAAGGGAGGCGCAGAGGTTTGGTGGGTGGGAAAGATTCT 34013  
QY 205 ----- 205  
Db 34012 TTTCTCCATTCTGATTAGTTTACCTGTCTCTGATTCAAAAGGCCAGGAAATTCCAATAGT 33953  
QY 205 ----- 205  
Db 33952 GAGGTAGGCTATTGGCTCTAGCAATTTCTTAATTCCTGTTTCCCTGTAGAGGTGGGGTGA 33893  
QY 205 ----- 205  
Db 33892 CCTGCATCGGAAAGAACTTGGCTTTCTATATGCTGTGCTCTGTTAAATTCACCTC 33833  
QY 205 ----- 205  
Db 33832 TGTCACTCTTATTCTCTCACTTGAGCAAGTTGCTTTACTTTTCCAAAGCCAGTTTCT 33773  
QY 205 ----- 205  
Db 33772 CATGAGAACTTTGTTTCAGGATTGTTGTTAACTGACTGAGCACTGTTAGTGCCATG 33713  
QY 205 ----- 205  
Db 33712 CATTTGTTCTTAGAGTATTGTGGATATCAACTAGTTCAATCTGTGTAACAACAGGTGAGG 33653  
QY 205 ----- 205  
Db 33652 ATAGATACTGTTATTTCAGTCTTACAGAAAAAGAAATTGAGGTATAGAGTTACTAAGTA 33593  
QY 205 ----- 205  
Db 33592 ACTTGCTTAGTAACCAAGTAACCTTAAGTAAGTAACTTAGTAACCCAGGTACTTACTTA 33533  
QY 205 ----- 205  
Db 33532 CTTGGTGCATACATGAGGAGCTGAGTTTGAACCTTTGAGAGTCTGACTCCAGCACTTGA 33473  
QY 205 ----- 205  
Db 33472 ACCTACTAAGCATTTGTACCTTTAATGTGTAAGAAAGTACCTGGCCTATCCTGGGCCCTCC 33413  
QY 205 ----- 205  
Db 33412 AGTGTGTTCTGTACCACCTCCTCTGTCTGTTCTTCCCTTAAGCCCAAGAAATCCGAA 33353  
QY 205 ----- 205  
Db 33352 CATCCAGATGCTGGGCCCTTGCCCTCACGGTCCAGGGCTAGGAGTTGAGTTCTTCCACTTA 33293  
QY 205 ----- 205  
Db 33292 GCTTCTTGCCCTGCTCCTCCTGCTATCTATCTAGCTGACCTCCAGCCCTGCCTCCTTAG 33233

QY 206 ValAspIlePheAsnLysGluLeuLeuLeuLeuProIleHisLeuGluValHisIrpSer 225  
Db 33232 GTGGACATCTTCAATAAAGAGCTCCTGCTAATCCCATCCACCTGGAGGTGCTATGGTCC 33173  
QY 226 LeuIleSerValAspValArgArgThrIleThrTyPheAspSerGlnArgThrLeu 245  
Db 33172 CTCACTCTGTTGAGCTGAGGCGGCGCACCATCACCTATTTTACTCGCAGGCGACCTG 33113  
QY 246 AsnArgArgCysProLys----- 251  
Db 33112 AACCCCGCTGCCCTAAAGTTTTCAGAGGGGAGGAAAGATGGCAGAAATGTGGGAGGAG 33053  
QY 251 ----- 251  
Db 33052 ATCCAGTGGCAAGGCATTTTCAGGCACAGAGTTGGGTTTGGGTCCTCTGATGAGCACCTG 32993  
QY 251 ----- 251  
Db 32992 GGCTTGGTGCCTGTGACCAGCACAGCTTGGTTTCAGTTGGGAAATGCAGAGGCTCACTTC 32933  
QY 252 -----HisIleAlaLysTyLeuGlnAlaGluAlaValLysLy 264  
Db 32932 CTTCTTCTCCTCATCATACATATTCCTCAAGTATTCAGGCGAGAGGAGTGAAGAA 32873  
QY 264 sAspArgLeuAspPheHisGlnGlyTyrPheLysMet----- 279  
Db 32872 AGACCGGCTGGATTTCCACCAGGCTGGAAGGTACTTCAAAATGGTAAGTTCTCTAGAG 32813  
QY 279 ----- 279  
Db 32812 GAGGTTTAAAGCAGGTTGAGTGGTGGAGGTGTGCTAGGAGCAGAAAGCTGAAGCTATA 32753  
QY 280 -----AsnValAlaArg- 283  
Db 32752 CACATAGGCTCTCCAGGAGAGGCTTACCTTCTTCTTCCAGAAATGGCCAGGG 32693  
QY 284 GlnAsnAspSerAspCysGlyAlaPheValLeuGlnTyPheCysLysHis 300  
Db 32692 CAGATAATGACAGTACTGTGGCCCTTGTGTGTCACGCAAGCAGATAC 32642  
RESULT 11  
AK000923  
LOCUS Homo sapiens cDNA FLJ10061 fis, clone HEMBA1001413.  
DEFINITION AK000923  
ACCESSION AK000923  
VERSION AK000923.1 GI:7021892  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to mRNA, clone lib:HEMBAL clone:HEMBAL1001413.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1  
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,  
Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H.,  
Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K.,  
Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,  
Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuno, Y. and Oshima, A.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 1578)  
Isogai, T. and Otsuki, T.  
Direct Submission  
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing and clone selection;  
Helix Research Institute (supported by Japan key technology Center  
etc.) and Department of Virology, Institute of Medical Science,  
University of Tokyo.

```

FEATURES
  source          Location/Qualifiers
1..1578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1001413"
/tissue_type="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/notes="cloning vector: pME18SFL3"
44..619
CDS
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAA91428.1"
/db_xref="gi:7021893"
translation="MSRNSLSIPVESLGHVYLMMLGSPFLGVGLTVLSDASVYPCGSL
ICMHSTWICPLCPQASLSCPPEPCPIGLLSCTLPNGFGQSGPGEERSLAPDGL
ILISNVCSIGHVAQELFOGSDGLGMAEAPERPKAGQHSPLREHVTVCVSIIDLEFL
QTYGLSLPLSTDEVEKLEIDIFQEEFFSPSR"
BASE COUNT      339 a 372 c 439 g 428 t
ORIGIN
Alignment Scores:
Pred. No.:      4,25e-72      Length:      1578
Score:          820.00      Matches:      186
Percent Similarity: 50.00%      Conservative: 10
Best Local Similarity: 47.45%      Mismatches: 17
Query Match:     46.28%      Indels:      179
DB:              9          Gaps:         4

US-09-848-852A-3 (1-332) x AK000923 (1-1578)
QY 1 MetTyrSerAlaGlnArgPheThrGlyThrIleTpaAlaArgArgGlyAlaHis-LeuAl 20
DB 274 ATGTACTCTGCCACAGGTTTGGGGGACATCTCGGCACAGAGGGGACCGAGCTTGGC 333
QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG1 40
DB 334 ACCCCCTGATGCCAGCATCTCATCAGCAATGTGTGCAGCATCGGGGACCATGTGGCCCA 393
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60
DB 394 GGAGCTTTTTCAGGCTCAGATTTGGGATGGCAGAGAGGAGGAGGAGGAGGAGGAGGAGAA 453
QY 60 sAlaGlyGlnHisSerProLeuArgGluHisValThrCysValGlnSerIleLeuAs 80
DB 454 AGCCGCGCCAGCACACCCCTCGAGAGAGGAGCATGTGACCTCGGTACAGAGCATCTTGA 513
QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLy 100
DB 514 CGAATTCCTTCAAGCTATGGCAGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGAA 573
QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSer----- 113
DB 574 GCTGGAGGACATTTTCCAGCAGGAGTTCACCCCTTCCAGGTCAGGCTTGAAGCCCT 633
QY 113 ----- 113
DB 634 CTTTGAAGAGAGGGCTGGGGCTTGGGATGTGGAGAGAACTACTGCTGCTTTCTTCCA 693
QY 113 ----- 113
DB 694 TAGGGCTGTAGTTGGGAGGAGGAAGCTAGACGTGAAGGGGAGGACTCTGACGGCAGGTG 753
QY 113 ----- 113
DB 754 CCAATTCCTTGAAGCTGATGGGAGAGTCTTTACCTGGGACCCCTGAAATGTTCTACCTGA 813
QY 113 ----- 113
DB 814 GTAGTCATGTTTATCTTCTGGGGAGTGGGCTTTTCGCGAGGTCTTCAGAGACCCATCA 873
QY 113 ----- 113

```

```

Db 874 TGAGCCAGAAAAAGGGAGTCAAGTTAGATTGTTATTCAGGGAGTAGTAGTATTCTG 933
QY 114 -----ArgLysGlyLe 117
Db 934 TGTGCCCCAGCTGCATCATCTTTTGTGTGACTCCACCCTTGGCCCTACTCAGGAAGGGCCT 993
QY 117 uValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPh 137
Db 994 GGTGTTGCAGCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGGTGGGGCTT 1053
QY 137 eArgValAlaTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGl 157
Db 1054 CGAGTGGCTTATAAGCGGCACGTGCTGACCATGTAGTACTTGGGACCTTGTATGGACA 1113
QY 157 nAsnTrpLeuAsnAsp----- 162
Db 1114 GAACCTGGCTCAGTGACCGAGGTGAGAAAGGTTAGAGAACAGGCCCTGAGAGGGGATTCCAG 1173
QY 162 ----- 162
Db 1174 GAGCAGGTGTCTGGGGCCCTCTGCATGGGGAGGCCCTGTACCCATGCCGACCCTCCAT 1233
QY 163 -----GlnValMetAsnMetTyrGlyAspLeuValMetAs 174
Db 1234 GGCAAGCTCCCTCCCATCTTCTCCCGAGGTGATGAACATGTATGGAGACCTGGTCATGA 1293
QY 174 pThrValProGluLysVal-----HisPhePheAsnSe 185
Db 1294 CACAGTCCCTGAAAGGTAGGCGCCCAACCAGATAGTCTACCCAGAGAACTTCTGCAG 1353
QY 185 rPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTyrThrLysAs 205
Db 1354 TTTT-----AGCAGCTTTTAAAGCTCTTTCATCTCTTTCATTATACACAGAGAG 1404
QY 205 nValAspIlePheAsnLysGluLeuLeuLeu 216
Db 1405 GGTCTCTGTTTTCAGGAGAGAGGTGTAGTGTA 1438
RESULT 12
AC099294 257644 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-39c15, *** SEQUENCING IN PROGRESS
DEFINITION ***, 71 unordered pieces.
ACCESSION AC099294
VERSION AC099294.4 GI:21731823
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 257644)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

```

Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,  
 Lozadó, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhiney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Picken, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,  
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmari, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstock, G., and Gibbs, R.  
 Direct Submission  
 2 (bases 1 to 257644)  
 Worley, K.C.  
 Direct Submission  
 Submitted (09-NOV-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 257644)  
 Worley, K.C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 11, 2002 this sequence version replaced gi:20258090.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GUSH  
 Center clone name: CH230-39C15  
 ----- Summary Statistics  
 Sequencing vector: plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 192798 bases at least Q40  
 Consensus quality: 198124 bases at least Q30  
 Consensus quality: 202420 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 71 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1665: contig of 1665 bp in length  
 \* 1666 1765: gap of unknown length  
 \* 1766 2950: contig of 1185 bp in length  
 \* 2951 3050: gap of unknown length  
 \* 3051 4076: contig of 1026 bp in length  
 \* 4077 4177: gap of unknown length  
 \* 4177 5352: contig of 1176 bp in length  
 \* 5353 5453: gap of unknown length  
 \* 5453 6762: contig of 1310 bp in length  
 \* 6763 6862: gap of unknown length  
 \* 6863 8152: contig of 1290 bp in length  
 \* 8153 8252: gap of unknown length  
 \* 8253 9323: contig of 1071 bp in length

9324 9423: gap of unknown length  
 9424 10907: contig of 1484 bp in length  
 10908 12554: gap of unknown length  
 12555 12554: contig of 1347 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1718 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1521 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1230 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1141 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1439 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1253 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1075 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1533 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1640 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1787 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1651 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1178 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 1671 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 2707 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 2137 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 2094 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 3247 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 3276 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 2411 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 2508 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 2970 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 3304 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 3066 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 2270 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 2545 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 2581 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 4242 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 3697 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 3012 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 2605 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 3715 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 3369 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 4528 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 3083 bp in length  
 12555 12554: gap of unknown length  
 12555 12554: contig of 98058: gap of unknown length

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## JOURNAL

## COMMENT



	* 98059	101384:	contig of 3326 bp in length	
	* 101385	101484:	gap of unknown length	
	* 101485	105449:	contig of 3965 bp in length	
	* 105450	105549:	gap of unknown length	
	* 105550	109216:	contig of 3667 bp in length	
	* 109217	109316:	gap of unknown length	
	* 109317	113575:	contig of 4259 bp in length	
	* 113576	113675:	gap of unknown length	
	* 113676	116968:	contig of 3293 bp in length	
	* 116969	117068:	gap of unknown length	
	* 117069	120956:	contig of 3888 bp in length	
	* 120957	121056:	gap of unknown length	
	* 121057	126598:	contig of 5642 bp in length	
	* 126599	126798:	gap of unknown length	
	* 126799	131591:	contig of 4793 bp in length	
	* 131592	131691:	gap of unknown length	
	* 131592	135524:	contig of 3833 bp in length	
	* 135525	135624:	gap of unknown length	
	* 135625	139656:	contig of 4032 bp in length	
	* 139657	139756:	gap of unknown length	
	* 139757	144725:	contig of 4969 bp in length	
Alignment Scores:				
Pred. No.:	2,29e-66	Length:	257644	
Score:	794.00	Matches:	164	
Percent Similarity:	62.43%	Conservative:	62	
Best Local Similarity:	45.30%	Mismatches:	89	
Query Match:	44.81%	Indels:	48	
DB:	2	Gaps:	6	
US-09-848-852A-3 (1-332) x AC099294 (1-257644)				
QY	11	IleTIPAlaArgGlyAlaHisLeuAlaProAspAlaSerIleLeuIleSer---	29	
Db	205854	CTCTGGGCAGGAACCGGCACAAGT-GCTGTCCAGCATGACGGAGTGAAAGTCGTCAATG	205912	
QY	30	-----AsnValCysSerIleGlyAspHisValalaGlnLeuPheGln	44	
Db	205913	TCTGGAGCAGACGAATCTGTGAGTAGTAGATGGCCCTGTGTCTGAGGAGCCGCTCAG	205972	
QY	45	GlySeraspLeuGlyMetAlaGlu---	52	
Db	205973	AATAAGAACTTCCAGATGGAGGAGGATGGTCTCTCAGGCAGAACATTCTTTAGTTTGA	206032	
QY	52	-----	52	
Db	206033	CTGCTGGACCACCCCTTACTCTAGACGTCACGTCGATGCCCTTGGTGTGTCAGTGAGCTC	206092	
QY	53	-----GluAlaGluArgProGlyGluLysAlaGlyGln---HisSerProLeuArgGlu	69	
Db	206093	AAAGTAGAAAAACCAAGTGGCAGGTGGCAAGTAGTGGCCAGACAGACGCTCCCCGTGGATGAT	206152	
QY	70	GluHisVal--ThrcysValGlnSerIleLeuAspGluPheLeuGlnThrTyrglySer	88	
Db	206153	GAGCAGCTCTCGACCTCGCTTCTGGTTCCTAGACGAGGTATGAAAAGTATGGAAGT	206212	
QY	89	LeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGlnGlu	108	
Db	206213	CTGCTCCACTCAGTGAGAAGATGTCCTTGGGAGATTAAAGATGCTCTTAATGAAGAC	206272	
QY	109	PheSerThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrglnArgMet	128	
Db	206273	TTTTCT-----AATAGAAAACCATTTATCAATFAGGAAATAACAACATACCGGGCCAGA	206326	
QY	129	ProGlyAsnAlaMetValargGlyPheArgValalaTyrllysArghisValLeuthrMet	148	
Db	206327	CATCAAAAGTGT-----AACTTCCGAATCTTCAACAAGCACATGCTGGATATG	206377	
QY	149	AspAspLeucglyThrLeuTyrglyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyr	168	
Db	206378	GACGACCTGCACACTTTGGATGTCAGAAATGTTGTAATGACCAAGTATTAATATGAT	206437	
QY	169	GlyaspLeuValMetAspThrvalProGluLysValHisPhePheAsnSerPhePheTyr	188	

[illegible]

the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method (Sugano et al., , Institute of Medical Science, University of Tokyo).

Custom primer used for sequencing  
( 5' end primer [CTTCTGCTCTAAAGCTGG];  
3' end primer [CGACCTCGAGCTCGACACA] ).

FEATURES

source  
1. .2007  
/organism="Macaca fascicularis"  
/db\_xref="taxon:9541"  
/clone="QtrA-13616"  
/sex="male"  
/tissue\_type="temporal lobe right"  
/clone\_lib="macaque brain cDNA library QtrA"  
/dev\_stage="adult"  
1140. .1442  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAB46895.1"  
/db\_xref="GI:13874562"  
/translation="WPGNAMVRGFRVAYKRHVLTMDDLGLYCGQWLNDOVRKRGRTG  
LRGDSGSRVSGALCHWGPCTHAAPSMASCLPSPQVMNMGDLVMDTVPEKVGPIR"  
BASE COUNT 469 a 465 c 551 g 522 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4, 65e-69 Length: 2007  
Score: 791.00 Matches: 177  
Percent Similarity: 50.86% Conservative: 1  
Best Local Similarity: 50.57% Mismatches: 2  
Query Match: 44.64% Indels: 170  
DB: 9 Gaps: 2

US-09-848-852A-3 (1-332) x AB060892 (1-2007)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20  
DB 382 GTGTACTCTGCCAACGGTTTGGGGACCATTCTGGGCCAAGAGGGGAGCGAGCTTGGC 441  
QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40  
DB 442 ACCCCCTGTATGCCAGCATCTCATCAACAAATGTGTGCAGCATCGGGACCATTGTGCCCA 501  
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60  
DB 502 GGAGCTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGGAGGAGGCTGGGGAGAA 561  
QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80  
DB 562 AGCCGGCCAGCAGACCCCTCGCAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGA 621  
QY 80 pGluPheLeuGlnThrTyrGlySerIleLeuProLeuSerThrAspGluValValGluLy 100  
DB 622 CGAATTCCTTCAACATATATGCGACGCTCATACCCCTCAGCAGCTGATGAGGTAGTGAGAA 681  
QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSer- 113  
DB 682 ATTTGGAGGACATTTTCCAGCAGGAGTTTCCACGCCTTCCAGGTGAGGCTTGAAGCCCT 741  
QY 113 ----- 113  
DB 742 CCTTCAAAGGAGGCTGGGGCCTCAGGAGGTAGGAGCTAGAGAGAAATACTCTGCTT 801  
QY 113 ----- 113  
DB 802 TTCTCCATAGGCTGTACTTTGGGGAGGAGGAAGCTAGAGCTGAAGGGGAGGACTCTGC 861  
QY 113 ----- 113  
DB 862 AGGCAAGGTGCCAGTCTTTGGAGCTGATGGAGAGTCTTTTACCTGGGACCCTGACATG 921  
QY 113 ----- 113

Db 922 TTCTACTTGAGTAGTATCATGTTCTTCTTGGGCACTGGGCCTTTTCGGGGTCTCTCAGAAGGA 981  
QY 113 ----- 113  
Db 982 CCCATCATGAGCCGAATAAAAGGGGTGCTAGTATTAGAAATTTGTATTCTAGGAGTAGTAGGT 1041  
QY 114 -----ArgL 115  
Db 1042 ATTTCTGTGCCCCAGCTGCATCAACTTTTGTGTGACTCCACCCTTGGCCTACTCAGGA 1101  
QY 115 ysglyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetVala 135  
Db 1102 AGAGCCTGGTGTTCAGCTGATCCAGTCATACCAAGCGATGCCAGCAATGCCATGGTGA 1161  
QY 135 rgGlyPheArgValAlaIaTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuT 155  
Db 1162 GGGCTTCCGTGTGGCTATAAGCGCATGTGCTGACCATGGATGATCTTTGGGACCTTGT 1221  
QY 155 yTgGlyGlnAsnTrpLeuAsnAsp- 162  
Db 1222 ATGGACAGAACTGGCTCATGACCAGGTGAGAAAGGGTAGAGNAACAGCGCCTGAGAGGG 1281  
QY 162 ----- 162  
Db 1282 ATTCAGGAGAGAGGGTGTCTGCGGCGCTCTGCAATGGGGAGCGCCTGTACCCATGCCGCAC 1341  
QY 163 -----GlnValMetAsnMetTyrGlyAspLeuV 172  
Db 1342 CCTCCATGGCAAGCTGCCCTCCCATCTTCTCCCGAGGTGATGAACATGTATGGAGACCTGG 1401  
QY 172 aMetAspThrValProGluLysVal 180  
Db 1402 TCATGGACACAGTCCCTGAAAGGTA 1427

RESULT 14

BC008589  
LOCUS Homo sapiens, clone IMAGE:3448367, mRNA, partial cds.  
DEFINITION BC008589  
ACCESSION BC008589  
VERSION BC008589.1 GI:14250318  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1830)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov  
COMMENT Contact: MGC help desk  
Email: cgabbs-r@mail.nih.gov  
Tissue procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
Sequencing Center  
Center code: BCM-HGSC  
Web site: http://www.hgsc.bcm.tmc.edu/cdna/  
Contact: villalona@bcm.tmc.edu.  
Villalona, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,  
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
Muzny, D.M., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: IRAC Plate: 19 Row: e Column: 11  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF



GRPVYWRQRLRHFGQRKALQIQKTIWIDHEILCAKTEKNAQTQWSTLSSKVRKQAK  
VHTSSKTLRLQAEKLLSSAKNSDHYETREKNLKLAVDTYPPNSALQOANGHRPTD  
QPSDFPMKNGSGSGSGITVYTLNNHKRGYCGCCGPHHRNGGLPIKFKFO  
LNOHRHTLSPILMAYEKLKSWIBERYLLRPSOHPHTKSVKCLRAQXSWQVKTQGOH  
ETRENGEGGSCFPSPSPKPDQSCRQYFPDMDSVAHWKGTNSHPDCHTTGSGSEL  
GKELSDDEAPQDOONGSANTDMQSCSDQPEWTHLIDTLPHERSNTLITSETER  
EIMTLQENQTSVSDRKLKLVSGADATSVSDGQPSQKAVONENYSQMEEDGSLKQ  
SILSLELHDHPYCKSPLEAPLVCSGKLTENQVGGKNSQKAVDDQLQSLSGLGF  
EVMKKYGSVLPSEKEVGLRGLKOVNDEFSNRKPFINREITNYARHQKCNFRIFYNK  
HMLDMDLATLQDQGNWLDQNMVGLMIDKAPDKVHFNSTFQKTVGTNGYKNG  
WTKVDLFLKSLLLIPIHELVHMSLITVLSNRIITSVDSOGIHFKFCVENIRKYLIT  
EAREKNRPEPLQGWQATVTKCIPOQKNDSDGCVFVLQCKCLALEQFQFQSDMDPRV  
PRTYKTLPCFRLMD\*

BASE COUNT	841 a	572 c	664 g	716 t
ORIGIN				

Alignment Scores:

Pred. No.:	1.18e-68	Length:	2793
Score:	789.00	Matches:	148
Percent Similarity:	73.72%	Conservative:	54
Best Local Similarity:	54.01%	Mismatches:	66
Query Match:	44.53%	Indels:	6
DB:	9	Gaps:	3

US-09-848-852A-3 (1-332) x BC030705 (1-2793)

Qy	58	GlyGluLysAlaGlyGlnHisSerProLeuArgGluGluHisValThr---CysValGln	76
Db	1656	GAAAGAACAGCTCAGAAAGCCCTCCAGCTGGATGATGAACAGCTGCTCAGTCTGCTTTCT	1715
Qy	77	SerIleLeuAspGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGlu	96
Db	1716	GGATTCTCTAGATGAGGTTATGAAGAAGTATGCACGTGTGGTTCACCTCAGTGA AAAAGAA	1775
Qy	97	ValValGluLysLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGly	116
Db	1776	GTCTCTGGAGATTAAGACATGCTTTAATGAAGACTTTCT-----AATAGAAAACCA	1829
Qy	117	LeuValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGly	136
Db	1830	TTTTCATTAATGAGGAATAACAACATATCGGGCCACACATCAAAATGT-----AAC	1880
Qy	137	PheArgValAlaTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGly	156
Db	1881	TTCCGTATCTTCTATAAATAAACACATGCTGGATATGGACGACCTGGCGACTCTGGATGT	1940
Qy	157	GlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrVal	176
Db	1941	CAGAACTGGCTGAATGACACGAGTCAATATATGTATGTGTAGCTGATGAATGGATGGACGT	2000
Qy	177	ProGluLysValHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyr	196
Db	2001	CCAGACAAAGTTCACCTCTCTCAACACCTTTTTCATACACAGCTGGTAACCAAGGATAT	2060
Qy	197	AspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIle	216
Db	2061	AATGGAGTAAAAAGATGGACTAAAAAGGTGGATTGTTTAAAAAGAGCTCTCTGTTGATT	2120
Qy	217	ProIleHisLeuGluValHisTrpSerLeuIleSerValAspValArgArgThrIle	236
Db	2121	CCTATTTCACCTCGAAGTCCACGTGGTCTCTCATATTACTGTGTGACACTCTCTAATCGAA	2180
Qy	237	ThrTyrPheAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyr	256
Db	2181	TCATTTTATGATTCCCAGGCATTTCATTTTAAAGTTTTTGTAGAGAATATAAGAAAGAT	2240
Qy	257	LeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyr	276
Db	2241	TTGCTGACTGAGCCACAGAAAAAATAAGACCTGNAATTTCTCAGGGTGGCAGACTGCT	2300
Qy	277	PheLysMetAsnValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGln	296
Db	2301	GTTACGAAGTGTATCCCAACAGAAAAACACACAGTCACTGTGGAGTCTTTGTGCTCCAG	2360

```
Qy   297 TyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAsnMetProLys    316  
      ||||| | ||||||| ||||| : ||||| : ||||| :  
Db   2361 TACTGCAAGTGCCTCGCCCTTAGAGCAGCCTTTTCACAGTATTTACAAGAAGACATGCCCGA    2420  
  
Qy   317 LeuArgargGlnIleTyrllysGluLeuCysHIsCysLysLeu     330  
      ::::|:::::|:|||:|||:||| |||::|:  
Db   2421 GTGGGAAGAGGATTACAAGGAGGTATGTGAGTGCCGCCTC    2462
```

Search completed: December 16, 2002, 22:23:25  
Job time : 3338 secs

Search completed: December 16, 2002, 22:23:25  
Job time : 3338 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2002, 19:39:15 ; Search time 46 Seconds  
(without alignments)  
2213.406 Million cell updates/sec

Title: US-09-848-852A-3  
Perfect score: 1772  
Sequence: 1 MYSQRFWTGTTIARRGAHLA.....DMPKLRRIQYKELCHCKLTV 332

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2\_1/USPto1/US0948852/runat\_16122002\_132037\_19035/app\_query.fasta\_1.519  
-DB=Issued\_Patents\_NA -QMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0948852 -CGCN\_1\_1\_38 @runat\_16122002\_132037\_19035 -NCFU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THRRADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	
1	108	6.1	4835	4	US-08-976-259-58	
2	93	5.2	1983	1	US-08-375-709-2	
3	93	5.2	1983	1	US-08-752-929-2	
c	4	93	5.2	9840	1	US-09-534-638-1
5	93	5.2	37895	1	US-08-375-709-1	
6	93	5.2	37895	1	US-08-752-929-1	
7	93	5.2	37895	3	US-09-090-793-1	
c	8	89.5	5.1	29793	4	US-09-302-812-38
c	9	89.5	5.1	29793	4	US-09-511-477-38
c	10	89.5	5.1	29793	4	US-09-511-507-38
11	88.5	5.0	43676	3	US-09-356-952-12	
12	87.5	4.9	3456	4	US-09-134-001C-405	
					Sequence 58, Appl	
					Sequence 2, Appl	
					Sequence 2, Appl	
					Sequence 1, Appl	
					Sequence 1, Appl	
					Sequence 1, Appl	
					Sequence 1, Appl	
					Sequence 38, Appl	
					Sequence 38, Appl	
					Sequence 38, Appl	
					Sequence 12, Appl	
					Sequence 405, App	

13	86	4.9	2184	4	US-09-134-001C-1230	Sequence 1230, Ap
14	84	4.7	925	2	US-08-531-927B-8	Sequence 8, Appli
15	84	4.7	5342	3	US-08-902-632-1	Sequence 1, Appli
16	84	4.7	5342	3	US-09-073-354-5	Sequence 5, Appli
17	84	4.7	5342	3	US-08-656-005A-5	Sequence 5, Appli
18	84	4.7	5342	3	US-09-073-259-5	Sequence 5, Appli
19	84	4.7	5342	4	US-09-363-095-5	Sequence 5, Appli
20	84	4.7	5342	4	US-09-418-027-5	Sequence 5, Appli
21	83	4.7	903	1	US-08-469-665-1	Sequence 1, Appli
22	83	4.7	903	2	US-09-038-596-1	Sequence 1, Appli
23	83	4.7	903	5	PCT-US95-00421-1	Sequence 1, Appli
24	83	4.7	1776	2	US-08-531-927B-1	Sequence 1, Appli
25	83	4.7	1776	4	US-09-041-886-12	Sequence 12, Appli
26	82.5	4.7	2023	3	US-08-961-083-199	Sequence 199, App
27	82.5	4.7	32768	4	US-08-961-527-71	Sequence 71, Appl
28	82	4.6	2900	1	US-08-034-650-9	Sequence 9, Appli
29	82	4.6	2900	1	US-08-449-015-9	Sequence 9, Appli
30	82	4.6	5339	3	US-09-073-354-6	Sequence 6, Appli
31	82	4.6	5339	3	US-08-656-005A-6	Sequence 6, Appli
32	82	4.6	5339	3	US-09-073-259-6	Sequence 6, Appli
33	82	4.6	5339	4	US-09-363-095-6	Sequence 6, Appli
34	82	4.6	5339	4	US-09-418-027-6	Sequence 6, Appli
35	81.5	4.6	1902	3	US-09-041-991A-9	Sequence 9, Appli
36	81	4.6	2007	4	US-09-134-001C-1979	Sequence 1979, Ap
37	81	4.6	11873	2	US-08-970-269A-32	Sequence 32, Appl
38	81	4.6	11873	4	US-09-407-562-32	Sequence 32, Appl
39	81	4.6	11878	2	US-08-970-269A-31	Sequence 31, Appl
40	81	4.6	11878	4	US-09-407-562-31	Sequence 31, Appl
41	81	4.6	11883	2	US-08-970-269A-28	Sequence 28, Appl
42	81	4.6	11883	4	US-09-407-562-28	Sequence 28, Appl
43	80.5	4.5	1350	4	US-09-134-001C-2836	Sequence 2836, Ap
44	80	4.5	873	2	US-08-531-927B-6	Sequence 6, Appli
45	80	4.5	2395	4	US-08-961-527-205	Sequence 205, App

#### ALIGNMENTS

#### RESULT 1

US-08-976-259-58  
; Sequence 58, Application US/08976259  
; Patent No. 6316609  
; GENERAL INFORMATION:  
; APPLICANT: Dillion, Patrick J.  
; APPLICANT: Choi, Gil H.  
; APPLICANT: Welch, Rodney A.  
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli  
; Patent No. 6316609  
; NUMBER OF SEQUENCES: 142  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Ave, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/976,259  
; FILING DATE: Herewith  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540

## ; INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4835 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-976-259-58

## Alignment Scores:

Pred. No.: 0.00929 Length: 4835  
Score: 108.00 Matches: 59  
Percent Similarity: 31.69% Conservative: 31  
Best Local Similarity: 20.77% Mismatches: 86  
Query Match: 6.09% Indels: 108  
DB: 4 Gaps: 11

US-09-848-852a-3 (1-332) x US-08-976-259-58 (1-4835)

QY 42 LeuPheGlnGlySerAspLeuGluMetAlaGluAlaGluArgProGlyGluLysAla 61  
Db 1350 CTATTTCCCTCGCTCGCTAGTATTGTGACGATGTTCCGAACAACCGACTTGGTGAAT 1409  
QY 62 GlyClnHisSerProLeuArgGluGluHisVal-----ThrCysValGlnSerIle--- 78  
Db 1410 CAAAGCCACAAATACCTGCATATGAACATATTTAGCTAGCTGTCCAGAGACATAGAC 1469  
QY 79 LeuAspGluPheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluVal 98  
Db 1470 CTGCTGCATTTATAGTGTGTGATGATTTTATGTCACCTGAATCGGAT----- 1520  
QY 99 GluLysLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuVal 118  
Db 1521 ACCAATTGTGCAGATTTTCTGAAAGACTTTCAGGATGAGAGTGCACGCTATTGCA 1580  
QY 119 LeuGlnLeuIleGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheArg 138  
Db 1581 TTG----- 1583  
QY 139 ValAlaTyrlsArgHisValLeuThrMetAspLeuGlyThrLeuTyrglyGlnAsn 158  
Db 1584 -----AAT 1586  
QY 159 TrpLeuAsn----- 161  
Db 1587 TGGCAAAATTTTGGTCTAGTGTGAATGTTTCTCAAGAGGGTGGTATTGAACGT 1646  
QY 162 -----AspGlnValMetAsnMetTyrglyAspLeuValMetAspThrVal 176  
Db 1647 TTTACCTATCGTCCCGCAATCCTTTAAGCTTCATCATAACTTCAAAAGCGTGGTCAA 1706  
QY 177 ProGluLysValHisPhePheAsnSerPhePheTyrlsAspLeuArgThrLysGlyTyr 196  
Db 1707 CCGAGAGGATTACCGCTTTTCAATCCCATATGCTGATTTGGTATGCTGCTGATAT 1766  
QY 197 AspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuIle 216  
Db 1767 -----ATCGATGATCGTGGTGTGCTGATTTGATTCGTCAC 1799  
QY 217 ProIleHis-----LeuGluValHisTrpSerLeuIleSerValAspVal 231  
Db 1800 CCGAGGATGATTAATGGGGTACTGCTGAAGTACCTGGAGCGGTGTCAGGGTA----- 1853  
QY 232 ArgArgArgThrIleThrTyrlsPheAspSerGlnArgThrLeuAsnArgArgCysProLys 251  
Db 1854 -----AATCACTATGCGATTA 1871  
QY 252 His-----IleAlaLysTyrlsLeuGlnAlaGluAlaValLysLysAspArg 266  
Db 1872 TCATTGAGGAATCTTGTGGCAAGCATCTGCGTGGTAGTCTGCCATGCTGAATCGA 1931  
QY 267 LeuAspPheHisGlnGlyTrpLysGlyTyrlsPheLysMetAsnValAlaArgGlnAsnAsn 286  
Db 1932 GTAAAGCAT-----AAAGATTATTTCAAG-----GCACATGATCGTAAT 1970

QY 287 AspSerAspCys 290  
Db 1971 GATGAAGAGTGC 1982

## RESULT 2

US-08-375-709-2  
; Sequence 2, Application US/08375709  
; Patent No. 5683898  
; GENERAL INFORMATION:  
; APPLICANT: YAZAWA, Kazunaga  
; APPLICANT: YAMADA, Akiko  
; APPLICANT: KATO, Seishi  
; APPLICANT: KONDO, Kiyosi  
; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid  
; TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of  
; TITLE OF INVENTION: Eicosapentaenoic Acid  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/375,709  
; FILING DATE: 20-JAN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/178,251  
; FILING DATE: 14-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-147945  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WEGNER, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1983 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM  
ORGANISM: BP-1625)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1983  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 1..1983  
US-08-375-709-2

## Alignment Scores:

Pred. No.: 0.169 Length: 1983  
Score: 93.00 Matches: 75  
Percent Similarity: 32.96% Conservative: 43  
Best Local Similarity: 20.95% Mismatches: 117  
Query Match: 5.25% Indels: 123  
DB: 1 Gaps: 19

US-09-848-852A-3 (1-332) x US-08-375-709-2 (1-1983)

```
QY 9 GlyThrIleTrp-AlaArgArgGlyAlaHisLeuAlaProProAspAlaSerIleIleuIl 28
Db 537 GGACCACATTGGCGAGCTGGCGGTTCAGAGAGATGTCCTGATGCAAAAGTACCG 596
QY 28 eSerAsnValCysSerIleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAspLe 48
Db 597 CTCG-----GATAACATCACTAAAGAAATGTCGATGAGAACGTACT 638
QY 48 uGly-----MetAlaGluGluAlaGluArgProGlyGluLysAlaGlyGlnHi 64
Db 639 TGCCGGTAACGCCATGAGCGCGCGGCGCTTATCAATACGCGCAACACTGGGCAACA 698
QY 64 s-----SerProLeuArgGluGluHisValThrCy 74
Db 699 TGACCACGGTATGTTGATGCTGCCTAGTAAAGGTCTATCAAAAGGTGAATCACTTA 758
QY 74 sValGlnSerIleLeuAspGluPheLeuGlnThrTrpGlySerLeuIleProLeuSerTh 94
Db 759 CGTCGCCCA-----GACTACACCTTAAACAGTGAAGCAATGGGAACGCTGACG-- 810
QY 94 rAspGluValValGluLysLeuGluAspIlePheGlnGlnPheSerThr----- 111
Db 811 -----APTGTGCTAGAGATGGTGTATTATGATGCTCGGGCACCAGCAAGCTCA 860
QY 112 -----ProSerArgLysGlyLeuValLeuGlnLeuIleGlnSe 124
Db 861 GTCAGAAATGATCACTTATATTCCTCTAAAGGCGCTCTGGAGCGGAGCTT--AC 917
QY 124 rTyrGlnArgMetProGly----- 130
Db 918 CTATCAAGGTATGCACACATTTATACGCTGCGCGCGCTAAAGTACGTGATCGCCTCA 977
QY 131 -----AsnAlaMetValArgGlyPhe-----ArgValAlaIarY 141
Db 978 GTGGTCAAAAGATATCAACGAAATGATCAATGCCTTTGTGTCAGATGTCGAAGTCTGT 1037
QY 141 rLysArgHisVal-----LeuTh 147
Db 1038 TGCCCTCGCAGCTCGCCGAGTGGGGTAACACGCGATCAACAGATTTCTTACGCTTACA 1097
QY 147 rMetAspAspLeuGlyThrLeuTyrGlyGln-----AsnTrpLeuAsnAspGlnVal-- 164
Db 1098 GCGTCAATACTACGCGCTAGTCACAACTCAACCTTGAGACTTGCACAGATGGTGTGCG 1157
QY 165 -MetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysValHisPhePheAs 184
Db 1158 TATACAAGATATTGGCGATGCGATTCAAGACACGATTCAGAGTCTATC----- 1206
QY 184 nSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTrpThrLy 204
Db 1207 -----TACAAGACGTGGCATACCAATGGTTTACCACGCGACT-----TATAGCCA 1250
QY 204 sAsnValAsp-----IlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHi 223
Db 1251 TAACTTAAGCGGTTTATACAACTATCTA----- 1281
QY 223 sTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAspSerGlnAr 243
Db 1282 -----GGCTACTTCGATATGAACCC 1301
QY 243 gThrLeuAsnArgArgCysPro-----LysHisIleAlaLysTyrLeuGln----- 258
Db 1302 AGCCAACTTATATCGCTGCGCAACCAACGAAATCTCGCAAGTTTGTGCAATACATGGG 1361
QY 259 -----AlaGluAlaValLysAspLeuAspPheHisGlnGlyTrpLysGlyTy 276
Db 1362 CGGCCAGATCGCGCAATTAAGCGCTAAAGATGATTACGCTCAAGGTGAATACCGCTT 1421
QY 276 rPhe-----LysMetAsnValAlaArgGlnAsnAspSer 288
Db 288 -----LysMetAsnValAlaArgGlnAsnAspSer 288
```

Db 1422 TGTTGCAACGGCATTAATAAGTGGTGGTGGCGGCGAGCCAGAAATGACTCC 1473

RESULT 3

US-08-752-929-2

Sequence 2, Application US/08752929

Patent No. 5798259

GENERAL INFORMATION:

APPLICANT: YAZAWA, Kazunaga

APPLICANT: YAMADA, Akiko

APPLICANT: KATO, Seishi

APPLICANT: KONDO, Kiyosi

TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing

TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic

TITLE OF INVENTION: Acid

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08752,929

FILING DATE: 20-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/375,709

FILING DATE: 20-JAN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/178,251

FILING DATE: 14-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 4-147945

FILING DATE: 15-MAY-1992

ATTORNEY/AGENT INFORMATION:

NAME: WEGNER, Harold C.

REGISTRATION NUMBER: 25,258

REFERENCE/DOCKET NUMBER: 53466/150/AAROK

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1983 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM

ORGANISM: BP-1625)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1983

FEATURE:

NAME/KEY: mat\_peptide

LOCATION: 1..1983

US-08-752-929-2

Alignment Scores:

Pred. No.: 0.169

Score: 93.00

Percent Similarity: 32.96%

Best Local Similarity: 20.95%

Query Match: 5.25%

DB: 1

Length: 1983

Matches: 75

Conservative: 43

Mismatches: 117

Indels: 123

Gaps: 19

```
US-09-848-852A-3 (1-332) x US-08-752-929-2 (1-1983)
Qy 9 GlyThrIleTrpAlaArgArgGlyAlaHisLeuAlaProProAspAlaSerIleLeuI 28
Db 537 GGACCACTTTGGCGGAGCTCGCGGTTCAGAGAGATGTTCCCTGATGTCAAAGTCTACGG 596
Qy 28 eSerAsnValCysSerIleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAspLe 48
Db 597 CTCA-----GATACATCACTAAGAAATGTCGATGAGAACGTACT 638
Qy 48 uGly-----MetAlaGluAlaGluArgProGlyGlyLysAlaGlyGlnH 64
Db 639 TGCGGTAAACCCATGAGCGCGCGAGCTTATCAATACGCGCAACACAGTGGCAACA 698
Qy 64 s-----SerProLeuArgGluGluHisValThrCy 74
Db 699 TGACCAAGGTATTGTTGATGCTCGGTAGTAAAGTCTATCAAAAGGTGAATCACTTA 758
Qy 74 sValGlnSerIleLeuAspGluPheLeuGlnThrTyGlySerLeuIleProLeuSerTh 94
Db 759 CGTGCAGCCCA-----GACTACACCTTAACAGTGAAGCGCAATGGGAACGCTGAG-- 810
Qy 94 rAspGluValValGluLysLeuGluAspIlePheGlnGlnGluPheSerThr----- 111
Db 811 -----ATTGATGGTCTAGAGATGGTGTATTATGGATGCTCGGCGCACCGAAGCTGA 860
Qy 112 -----ProSerArgLysGlyLeuValLeuGlnLeuIleGlnSe 124
Db 861 GTCAGAAATGATCATTATATCCCTCTAAAAAGCGCTCTGGACGGCGAGCTT---AC 917
Qy 124 rTyGlnArgMetProGly----- 130
Db 918 CTATCAAGGTATCCACAACTTTATACGCTGCGCGCGCTAAAGTACGTGATCGGCTCAA 977
Qy 131 -----AsnAlaMetValArgGlyPhe-----ArgValAlaTy 141
Db 978 GTGCTCAAAAGATATCAACGAATATGATCAATGCTTTGGTCAAGATGTCGAAGTGTGT 1037
Qy 141 rLysArgHisVal-----LeuTh 147
Db 1038 TGCTCTGCACCTGCGCCAGTGTGGGTAAACAGCGATCAACGATTTCTTACGCGCTACA 1097
Qy 147 rMetAspAspLeuGlyThrLeuTyGlyGln-----AsnTrpLeuAsnAspGlnVal-- 164
Db 1098 GCGTGATACTACGCGCTAGTGCACATCAACCTTGAGACTTGCCACGATGGTGTGCG 1157
Qy 165 -MetAsnMetTyGlyAspLeuValMetAspThrValProGluLysValHisPhePheAs 184
Db 1158 TATACAAGATATTGGCGATCGGATTCAGACACAGATTCAGAGTCTATC-----1206
Qy 184 nSerPhePheTyAspLysLeuArgThrLysGlyTyAspGlyValLysArgTrpThrLy 204
Db 1207 -----TACAAGCGTGGCATACCAATGTTTACCACGCGACT-----TATAGCCA 1250
Qy 204 sAsnValAsp---IlePheAsnLysGluLeuLeuIleProIleHisLeuGluValH 223
Db 1251 TAACGCTAAAGCGGTTTATAACAAGTATCTA-----1281
Qy 223 sTrpSerLeuIleSerValAspValArgArgThrIleThrTyPheAspSerGlnAr 243
Db 1282 -----GGCTACTCTCGATATGAACCC 1301
Qy 243 gThrLeuAsnArgArgCysPro-----LysHisIleAlaLysTyLeuGln----- 258
Db 1302 AGCCAACTTAAATCCGCTGCCCAACCAAGCAAGATCTGCCAAGTTGTCCGAATACATGG 1361
Qy 259 -----AlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTy 276
Db 1362 CGCGGAGATGCGGCAATTAAGCGGCTAAAGATGATTACGCTCAAGGTGAATACCGCTT 1421
Qy 276 rPhe-----LysMetAsnValAlaAlaArgGlnAsnAspSer 288
Db -----LysHisIleAlaLysTyLeuGlnAlaGluAlaValLysLysAspArgLeu 267
```

Db 1422 TGTTCACGGCATTAAATAAGGTGGTGTGATGCGCGAGCCAGAAATCACTCC 1473

RESULT 4  
US-09-534-638-1/c  
; Sequence 1, Application US/09534638  
; Patent No. 6320038  
; GENERAL INFORMATION:  
; APPLICANT: Panula, Pertti A.J.  
; APPLICANT: Brandt, Annika  
; APPLICANT: Westerlund, Johanna  
; TITLE OF INVENTION: Promoter for Neuropeptide FF Promoter and use thereof  
; TITLE OF INVENTION: for therapy and diagnosis  
; FILE REFERENCE: 2530-104  
; CURRENT APPLICATION NUMBER: US/09/534,638  
; EARLIER APPLICATION NUMBER: 09/365755  
; EARLIER FILING DATE: 1999-08-03  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 9840  
; TYPE: DNA  
; ORGANISM: Mouse  
US-09-534-638-1

Alignment Scores:  
Pred. No.: 2,12 Length: 9840  
Score: 93.00 Matches: 48  
Percent Similarity: 40.80% Conservative: 34  
Best Local Similarity: 23.88% Mismatches: 67  
Query Match: 5.25% Indels: 52  
DB: 4 Gaps: 10

US-09-848-852A-3 (1-332) x US-09-534-638-1 (1-9840)

Qy 92 LeuSerThrAspGluValValGluLysLeuGluAspIlePhe-----105  
Db 3755 TTGGGAAAAGATGAGTTA-----AAACTGGAGGACATAGCCACTCTGACTGTGACAA 3702  
Qy 106 -----GlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGln 120  
Db 3701 AATTCACACAGAGAAACACAAAGAACATGAGACACAGCAGCGGCATACTA-----3648  
Qy 121 LeuIleGlnSerTyGlnArgMetPro-----GlyAsnAlaMetValArgGlyPheArg 138  
Db 3647 -----CAACCTCTCTTAATGAAAACCAAAAGTACTGAGAGTCTGAGA 3603  
Qy 139 ValAlaTyLys-----ArgHisValLeuThrMetAspLeuGlyThrLeuTy 155  
Db 3602 AAAGGTTCAAGTTGATGTAGAGTCTCAGATACAAATCGATGAACCTAGGAACCTGTATAC 3543  
Qy 156 GlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyGlyAspLeuValMetAspThr 175  
Db 3542 AAAACCTCCAAATGAACAAAACAGTTG-----GATAAG 3510  
Qy 176 ValProGluLysValHisPheAsnSerPhePheTyAspLysLeuArgThrLysGly 195  
Db 3509 TTGGAGGAGAAAGTCAGATATATGAGGAGAGATTTT-----3474  
Qy 196 TyrAspGlyValLysArgTrpThrLysAsn---ValAspIlePheAsnLysGlu-----212  
Db 3473 GATGAGGCGATTCAAACATGGAAAGGAAACAGATAGAAATTAACATGAAGAACAACATCA 3414  
Qy 213 -----LeuLeuLeuIleProIleHisLeuGluValHisTrpSerLeuIle 227  
Db 3413 AAGTGAACCCCTCACTAGAAAGTGTCCCAAGAGATTAAGATCATCGATGTTCT-----3360  
Qy 228 SerValAspValArgArgThrIleThrTyPheAspSerGlnArgThrLeuAsnArg 247  
Db 3359 GCAGTGGAGCACAAAGTTGGAAACAATCATATCATCAGACACACACAGACAGACAGACA 3300  
Qy 248 ArgCysProLysHisIleAlaLysTyLeuGlnAlaGluAlaValLysLysAspArgLeu 267



Db 3299 CAGACAGACACACACACACACACACAGACAGACAGAGAGAGAGAGAGAGAGAGATA 3240

Qy 268 Asp 268

Db 3239 GAT 3237

RESULT 5

US-08-375-709-1

; Sequence 1, Application US/08375709

; Patent No. 5683898

; GENERAL INFORMATION:

; APPLICANT: YAZAWA, Kazunaga

; APPLICANT: YAMADA, Akiko

; APPLICANT: KATO, Seishi

; APPLICANT: KONDO, Kiyosi

; TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid

; TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of

; TITLE OF INVENTION: Eicosapentaenoic Acid

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/375,709

; FILING DATE: 20-JAN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/178,251

; FILING DATE: 14-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 4-147945

; FILING DATE: 15-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: WEGNER, Harold C.

; REGISTRATION NUMBER: 25,258

; REFERENCE/DOCKET NUMBER: 53466/150/AAOK

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 37895 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:

; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM

; ORGANISM: BP-1625)

US-08-375-709-1

Alignment Scores:

Pred. No.: 17.8

Score: 93.00

Length: 37895

Matches: 75

Percent Similarity: 32.96%

Conservative: 43

Best Local Similarity: 20.95%

Mismatches: 117

Query Match: 5.25%

Indels: 123

Gaps: 19

US-09-848-852A-3 (1-332) x US-08-375-709-1 (1-37895)

Qy 9 GlyThrIleTrp-AlaArgArgGlyAlaHisLeuAlaProProAspAlaSerIleLeuI 28

||||| :|||

Db 6657 GGACCACATTGGCGAGCTCGCGGTTCACAGAGATGTTCCCTGATGTCAAGTCTACGG 6716

Qy 28 eSerAsnValCysSerIleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAspLeu 48

Db 6717 CTCA-----GATAACATCACTAAAGAAATTCGATGAGAACGTACT 6758

Qy 48 uGly-----MetAlaGluGluAlaGluArgProGlyGluLysAlaGlyGlnHi 64

Db 6759 TGCCCGTAAGCCATGAGCGCGCGCGAGCTTATCAATACGGCGCAACACTGGCGCAACA 6818

Qy 64 s-----SerProLeuArgGluGluHisValThrCy 74

Db 6819 TGACCACGGTATTGTTGATGTCGCTAGTAAAGGTATATCAAAAGTGAATACACTTA 6878

Qy 74 sValGlnSerIleLeuAspGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerTh 94

Db 6879 CGTCCCCCA-----GACTACACCTTAAACAGTCAAGGCAAAATGGGAACCTGACG-- 6930

Qy 94 rAspGluValValGluLysLeuGluAspIlePheGlnGlnGluPheSerThr----- 111

Db 6931 -----ATTGATGGTCTAGAGATGGTGTATTGATGCTCGGGCACCAGCAAGCTGA 6980

Qy 112 -----ProSerArgLysGlyLeuValLeuGlnLeuIleGlnSe 124

Db 6981 GTCAGAAATGATCACTTATATTCCTCTAAAAAAGCGCTCTGGACGGCGAGCTT---AC 7037

Qy 124 rTyrGlnArgMetProGly----- 130

Db 7038 CTATCAAGGTATGCACACACATTTATACGTCGCGCGCGCTAAAGTACGTGTCGCTCAA 7097

Qy 131 -----AsnAlaMetValArgGlyPhe-----ArgValAlaIaTy 141

Db 7098 GTGGTCAAAAGATATCAACGAAATGATCAATGCCCTTTGGTCAAGATGTGGAAGTCTGT 7157

Qy 141 rLysArgHisVal-----LeuTh 147

Db 7158 TGCCCTCGCACTCTGCGCCAGTGTGGGTAAACCAAGCATCAACGATTTCTTACGCTACA 7217

Qy 147 rMetAspAspLeuGlyThrLeuTyrGlyGln-----AsnTrpLeuAsnAspGlnVal-- 164

Db 7218 GCGTGATTAACCTAGCGCTAGTGCACAACTCAACCTTGAGACTTGCACAGATGGTGTGG 7277

Qy 165 -MetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysValHisPheAs 184

Db 7278 TATACAAGATATTGCGGATCGGATTCAGACACAGATTCAGAGTCTATC----- 7326

Qy 184 nSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTrpThrLy 204

Db 7327 -----TACAAGACGTGGCATACCAATGGTTTACCACGGCACT-----TATAGCCA 7370

Qy 204 sAsnValAsp---IlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHi 223

Db 7371 TAACGCTAAAGCGGTTTATACCAAGTATCTA----- 7401

Qy 223 sTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAspSerGlnAr 243

Db 7402 -----GGCTACTTCGATGAACCC 7421

Qy 243 gThrLeuAsnArgArgCysPro-----LysHisIleAlaLysTyrLeuGln----- 258

Db 7422 AGCCCAACCTTAATCCGCTGCCCAACCAAGCAAGAACTCTGCCAAGTTTGTGGAATACATG 7481

Qy 259 -----AlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTy 276

Db 7482 CGGCGCAGATGCGCAATTAAAGCGGCTAAAGATGATTACGCTCAAGGTGAATACCGCTT 7541

Qy 276 rPhe-----LysMetAsnValAlaArgGlnGlnAsnAspSer 288

Db 7542 TGTTCCACCGCATTAATAAGTGGTGTGTCGCCGCCGACCAAGAAATGACTCC 7593

RESULT 6

US-08-752-929-1

; Sequence 1, Application US/08752929

Patent No. 5798259  
GENERAL INFORMATION:  
APPLICANT: YAZAWA, Kazunaga  
APPLICANT: YAMADA, Akiko  
APPLICANT: KATO, Seishi  
APPLICANT: KONDO, Kiyoshi  
TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing  
Enzymes and Process for Production of Eicosapentaenoic  
Acid  
TITLE OF INVENTION: Acid  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,929  
FILING DATE: 20-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/375,709  
FILING DATE: 20-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/178,251  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-147945  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 37895 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM  
BP-1625)  
US-08-752-929-1

Alignment Scores:  
Pred. No.: 17, 8 Length: 37895  
Score: 93.00 Matches: 75  
Percent Similarity: 32.96% Conservative: 43  
Best Local Similarity: 20.95% Mismatches: 117  
Query Match: 5.25% Indels: 123  
DB: 1 Gaps: 19

US-09-848-852a-3 (1-332) x US-08-752-929-1 (1-37895)

Qy 9 GlyThrIleTrpAlaAargGlyAlaHisLeuAlaProAspAlaSerIleLeu1 28

Db 6657 GGACCACTTTGGCGGAGCTGCTCAAGAGATGTTCCCTGATGTCAAAAGTCTACGG 6716

Qy 28 eSerAsnValCysSerIleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAsp16 48

Db 6717 CTCA-----GATAACATCACTAAAGAAATTTGTCGATGAGAACCTACT 6758

Qy 48 uGly-----MetAlaGluGluAlaGluAargProGlyGluLysAlaGlyGlnHi 64  
Db 6759 TGCCCGTAACGCATCAGCGCGCGCAGCTTATCAATACGCGCAACACGTCGCAACA 6818  
Qy 64 s-----SerProLeuAargGluGluHisValThrCy 74  
Db 6819 TGACCACGCTATTGTTGATGCTCGCTAGTAAGTCTATCAAAAGGTGAATCACTTA 6878  
Qy 74 sValGlnSerIleLeuAspGluPheLeuGlnThrTyGlySerLeuIleProLeuSerTh 94  
Db 6879 CGTCGCCCA-----GACTACACCTTAAACAGTGAAGCAATGGAAACGCTGAGC-- 6930  
Qy 94 rAspGluValValGluLysLeuGluAspIlePheGlnGlnGluPheSerThr----- 111  
Db 6931 -----ATTGATGGTCTAGAGATGGTGTATTGATGGCTCGCGGACCGCAAGCTGA 6980  
Qy 112 -----ProSerArgLysGlyLeuValLeuValLeuGlnLeuIleGlnSe 124  
Db 6981 GTCAGAAATGATCATTATTCCTCTAAAAAAGCGCTCTGGACGGCGGAGCTT--AC 7037  
Qy 124 rTyrglnArgMetProGly----- 130  
Db 7038 CTATCAAGGTATGCCACAACATTTATACGCTGCCGCGCGCTAAAGTAGCTGATCGGCTCAA 7097  
Qy 131 -----AsnAlaMetValArgGlyPhe-----ArgValAlaIar 141  
Db 7098 GTGCTCAAAAGATATCAACGAAATGATCAATGCTTTGGTCAAGATGTCGAAGTGTGT 7157  
Qy 141 rLysArgHisVal-----LeuTh 147  
Db 7158 TGCTCGCACCTCTGCGCCAGTGTGGGTAACCAAGCGATCAACGATTTCTTACGCTACA 7217  
Qy 147 rMetAspAspLeuGlyThrLeuTyrgln-----AsnTrpLeuAsnAspGlnVal-- 164  
Db 7218 GCCTGATAACTACGGCCTAGTGCACAATCAACCTTGAGACTTGCACAGATCTATC----- 7326  
Qy 165 -MetAsnMetTyrglyAspLeuValMetAspThrValProGluLysValHisPhePheAs 184  
Db 7278 TATACAAGATATTGGCGATCGGATTCACAGACAGATTCAGAGATCTATC----- 7326  
Qy 184 nSerPhePheTyrglyAspLeuAargThrLysGlyTyrglyAspGlyValLysArgTrpThrLy 204  
Db 7327 -----TACAAGACGTGGCATACCAATGTTTACCACGCGACT-----TATAGCCA 7370  
Qy 204 sAsnValAsp---IlePheAsnLysGluLeuLeuIleProIleHisLeuGluValHi 223  
Db 7371 TACGCTAAAGCGGTTTATAACAAGTATCTA----- 7401  
Qy 223 sTrpSerLeuIleSerValAspValAargArgArgThrIleThrTyrglyPheAspSerGlnAr 243  
Db 7402 -----GGCTACTTCTGATATGAACCC 7421  
Qy 243 gThrLeuAsnAargCysPro-----LysHisIleAlaLysTyrglyGln----- 258  
Db 7422 AGCCAACTTAATCCGCTGCTCAACCAAGCAAGAAATCTGCCAAGTTTGTGGAATACATGGG 7481  
Qy 259 -----AlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTy 276  
Db 7482 CGGCGCAGATGCGCAATTAAGCGCGCTAAAGATGATTCGCTCAAGGTGAATACCGCTT 7541  
Qy 276 rPhe-----LysMetAsnValAlaAargGlnAsnAsnAspSer 288  
Db 7542 TGTTCGAACGGCATTAATAAGTGTGTCGCGCGGAGCCAGAAATGACTCC 7593

## RESULT 7

US-09-090-793-1  
; Sequence 1, Application US/09090793  
; Patent No. 6140486  
; GENERAL INFORMATION:  
; APPLICANT: Calgene, LLC  
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
; TITLE OF INVENTION: of polyketide-like synthesis genes in plants  
; FILE REFERENCE: CGNE.131.01US

Qy 135 ArgGlyPheArgValAlaTyrLysArgHisValLeuThrMetAspLeu----- 151

Db 26755 -----AAAATAAATGAGGATCTTCAAAAACAATTTTATGATTTGTGCCCATAT 26705  
QY 152 -----GlyThrLeuTyrGly-----GlnAsnTrpLeu 160  
Db 26704 CCGGTTAACGGTTACGGTGAATATTAATTTGGTGATGAACACCGAAACAGAGAGAGTTG 26645  
QY 161 AsnAspGlnValMetAsnMet-----TyrGlyAspLeu--- 171  
Db 26644 ATGGATCAACTTTTGGCAAAATTAAGATTAATAATGATGAGATTTATGACATTTTC 26585  
QY 172 ---ValMetAspThrValProGluLysValHisPhePheAsnSerPhePheTyrAsp--- 189  
Db 26584 CACTTGCTGGAACATATTGACAAGATCATAGTACAGTTTCTCTTATTTCAAAGGATCCT 26525  
QY 190 -----LysLeuArgThrLysGlyTyrAspGly 198  
Db 26524 ATACTTATACAAAAAGATGGAAGACAACACAGATCATTTCAAAAACCTGGAAGACGAA 26465  
QY 199 ValLysArgTrpThrLysAsnValAspPheAsnLysGluLeuLeuLeuProIle 218  
Db 26464 GTTGACTATCTTACGAAAACTTGGAAAGTTTGAATCCATCGCAGCTGGTTACATCTTTG 26405  
QY 219 HisLeuGluValHisTrpSerLeuLeuSerValAspVal---ArgArgArgThrIleThr 237  
Db 26404 GAAACAAAGATGATTCTCTTAGTGACAGTGGAAATTCCAAAAAGAAAGAAATATTCTG 26345  
QY 238 TyrPheAsp-SerGlnArgThrLeuAsnArg---ArgCysProLysHisIleAlaLysTy 256  
Db 26344 TTTTCGAACTCAAGAAAAGATGAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTT 26285  
QY 256 rLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTy 276  
Db 26284 TATCAGAAGGATAGTATTGAAAAG-----TCATGGAGAAATTT 26245  
QY 276 rPheLysMetAsn 280  
Db 26245 TTGGAACCTTCAAC 26233

RESULT 9  
US-09-511-477-38/c  
; Sequence 38, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 38  
; LENGTH: 29793  
; TYPE: DNA  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-511-477-38  
Alignment Scores:  
Pred. No.: 33.3 Length: 29793  
Score: 89.50 Matches: 54  
Percent Similarity: 41.51% Conservative: 56  
Best Local Similarity: 20.38% Mismatches: 81  
Query Match: 5.05% Indels: 74  
DB: 4 Gaps: 13  
US-09-848-852a-3 (1-332) x US-09-511-477-38 (1-29793)

QY 59 GluLysAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIle 78  
Db 26932 GAAAGAGATCAAAATATCAAAAATTTAAAGAGAA-----GTTGAGAAATTA 26885  
QY 79 LeuAspGluPheLeuGlnThrTyrGlySerLeuLeuProLeuSerThrAspGluValVal 98  
Db 26884 ATCGATGAA-----CAAGAAGATGAGATAATA 26858  
QY 99 -----GluLysLeuGluAspPheGlnGlnPheSerThrProSerArg----- 114  
Db 26857 CTAGCTGAAATCGAATTCGAAATTTTCCAAATAAAGACGACAGACAGCTAGAAATG 26798  
QY 115 LysGlyLeuValLeuGlnLeuLeuGlnSerTyrGlnArgMetProGlyAsnAlaMetVal 134  
Db 26797 AGAAGCTTGTGATATGAGAAATGAGATTTCTCAATCTC----- 26756  
QY 135 ArgGlyPheArgValAlaTyrLysArgHisValLeuThrMetAspAspLeu----- 151  
Db 26755 -----AAAATAAATGAGGATCTTCAAAAACAATTTTATGATTTGTGCCCATAT 26705  
QY 152 -----GlyThrLeuTyrGly-----GlnAsnTrpLeu 160  
Db 26704 CCGGTTAACGGTTACGGTGAATATTAATTTGGTGATGAACACCGAAACAGAGAGAGTTG 26645  
QY 161 AsnAspGlnValMetAsnMet-----TyrGlyAspLeu--- 171  
Db 26644 ATGGATCAACTTTTGGCAAAATTAAGATTAATAATGATGAGATTTATGACATTTTC 26585  
QY 172 ---ValMetAspThrValProGluLysValHisPhePheAsnSerPhePheTyrAsp--- 189  
Db 26584 CACTTGCTGGAACATATTGACAAGATCATAGTACAGTTTCTCTTATTTCAAAGGATCCT 26525  
QY 190 -----LysLeuArgThrLysGlyTyrAspGly 198  
Db 26524 ATACTTATACAAAAAGATGGAAGACAACACAGATCATTTCAAAAACCTGGAAGACGAA 26465  
QY 199 ValLysArgTrpThrLysAsnValAspPheAsnLysGluLeuLeuLeuProIle 218  
Db 26464 GTTGACTATCTTACGAAAACTTGGAAAGTTTGAATCCATCGCAGCTGGTTACATCTTTG 26405  
QY 219 HisLeuGluValHisTrpSerLeuLeuSerValAspVal---ArgArgArgThrIleThr 237  
Db 26404 GAAACAAAGATGATTCTCTTAGTGACAGTGGAAATTCCAAAAAGAAAGAAATATTCTG 26345  
QY 238 TyrPheAsp-SerGlnArgThrLeuAsnArg---ArgCysProLysHisIleAlaLysTy 256  
Db 26344 TTTTCGAACTCAAGAAAAGATGAAGATTAAGATTAAGATTAAGATTAAGATTT 26285  
QY 256 rLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTy 276  
Db 26284 TATCAGAAGGATAGTATTGAAAAG-----TCATGGAGAAATTT 26246  
QY 276 rPheLysMetAsn 280  
Db 26245 TTGGAACCTTCAAC 26233  
RESULT 10  
US-09-511-507-38/c  
; Sequence 38, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG)  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREAC  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 38  
; LENGTH: 29793  
; TYPE: DNA  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-511-507-38



```

QY 204 LysAsnValAspIlePheAsnLysGlnLeuLeuLeuProIleHisLeuGluValHis 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 32796 GTTAACTAGCAACATTAAAGAATATTTTTCAGCAACAACATTTTCATATA----- 32846

QY 224 TrpSerLeuIleSerValAspValArgArgThrIleThrTyPheAspSerGlnArg 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 32847 TGGAAAGCTA-----AGCATAAAGAATAATAACTACCACGGGTGGGAAGA 32891

QY 244 ThrLeuAsnArgCysProLysHisIleAlaLysTyLeuGlnAlaGluAlaValLys 263
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 32892 AGGATTTTCGAAGATAAACAGACNAGAGTTATAAATCATGTAATCAATAAGACATTCCA 32951

QY 264 LysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyPheLysMetAsnValalaArg 283
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 32952 GAAAAGTGAGGTCGTTTTCTCTACAAAGAAAAATATCTGAATAAATGGAAAAAGAAAAAC 33011

QY 284 GlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyCysLysHisLeu----- 301
    |||:||| |||:||| |||:||| |||:||| |||:||| |||:||| |||:|||
DB 33012 ATTGAAAACGAAGATAAACTTTGGGCACCTTATGAACCTGGAGAATAAATTCATCAAACAA 33071

QY 302 -----AlaLeuSerGlnProPheSerPheThrGlnGlnAsp 313
    |||:||| |||:||| |||:||| |||:||| |||:||| |||:||| |||:|||
DB 33072 AAGTTTTTCGCAAAATTAACCGGTCATTTCAACATAGTCAACAAGAG 33119

RESULT 12
US-09-134-001C-405
; Sequence 405, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAB
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 405
; LENGTH: 3456
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-405

Alignment Scores:
Pred. No.: 1.97 Length: 3456
Score: 87.50 Matches: 60
Percent Similarity: 43.36% Conservative: 51
Best Local Similarity: 23.44% Mismatches: 96
Query Watch: 4.94% Indels: 50
DB: 4 Gaps: 14

US-09-848-852A-3 (1-332) x US-09-134-001C-405 (1-3456)

QY 98 ValGlnLysLeuGluAspIlePheGlnGlnGlnPheSerThrProSerArgLysGlyLeu 117
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 2591 ATAAATGNATTAAAGAAAGTAGTATACATCTTCCAAAGGTTGAATCTAGACAATTTTA 2640

QY 118 ValLeuGlnLeuIleGlnSerTyrrGlnArgMetProGlyAsn-----AlaMetValArg 135
    |||:||| |||:||| |||:||| |||:||| |||:||| |||:||| |||:|||
DB 2641 GAGCAGCAGATGACTTATCAATTCCAAACAATGAATAACTCCITTTAAAAAATGATTAAA 2700

QY 136 GlyPhe-----ArgValalaTyrrLysArgHisValLeuThrMetAspasp 150
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 2701 GATTTTTCGATGACAAACGCCAAATTTATCAGCTAGACAACAT-----AAA 2745

QY 151 LeuGlyThrLeuTyrrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrrGlyAsp 170
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 2746 TTAATCAGCTATAAAAACCAACCCCTTAATTAATCAAGAGTTGAVTTGATGACAACTGCACA 2805

QY 171 LeuValMetAspThrValProGluLysValHisPheAsnSerPhePheTyrrAspLys 190

```

```

Db 2806 CGTACTTTCAATGAAGTAGAAGACAAGATATATCATCTAAAT---GAACGG 2853
      :::: ||| |||:::||||::: ||| |||
Qy 191 LeuArgThrLysGlyTyrAspGlyValLys-----ArgTyrThrLysAsn 205
      :::: :::: ||||| ||||| :::: |||:::||||
Db 2854 TTAACAACATACAACTTTAGATGAGGTTAAATCTGTGTTTAATAGTCAGATGACACAAAT 2913
      :::: :::: ||||| ||||| :::: |||:::||||
Qy 206 ValAspIlePheAsnLysGlu-----LeuLeuIlePheIleHis 219
      ||| |||||:::|||| :::: ||| |||||
Db 2914 AACGAC---TTTAATGAGCAAAAGCAAAATTTCACTAAATATATTTAGATCAAAATTCAT 2970
      ||| :::: ||||| ||||| :::: |||
Qy 220 -----LeuGluValHisTyrSerLeuIleSerValAspValArgArgThrIleThr 237
      ||| :::: ||||| ||||| :::: |||
Db 2971 CAACGCTTATCTTGTAGACCAATCACTTATCATCACAAAGGATTAAAAA----- 3018
      ||| :::: ||||| ||||| :::: |||
Qy 238 TyrPheAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeu 257
      |||||:::|||| ||| :::: |||
Db 3019 TATTTTAATTCACAA-----CTAGAAGAACAAATCTCACAGTCATGAAAGGTTAAAT 3072
      |||||:::|||| ||| :::: |||
Qy 258 GlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGly-Tyr-----LysGln 275
      ||| :::: ||| :::: |||
Db 3073 CAGATTCACTGTTATATATAAATGCAAAATTTAATGTTGAGCCATCACTGGTTGATACGGCG 3132
      |||||:::|||| ||| :::: |||
Qy 275 yTyrPheLysMetAsnValAlaArg-----GlnAsnAsnAspSerAspCys 290
      |||||:::|||| ||| :::: |||
Db 3133 TTACTTCAAAATGAACTTAATTCAAATGTTGCAATCACTACCAAAACAGTTAACTAAACGT 3192
      |||||:::|||| ||| :::: |||
Qy 290 sGlyAlaPheValLeuGlnTyrCysLys-----HisLeuAlaLeuSerGlnPr 306
      :::: :::: ||| :::: |||
Db 3193 AAAATAGTAACTCAAAAGTCACAA-AAGGATATACAAAGTACACATAGCT---AATCAAAC 3248
      :::: :::: ||| :::: |||
Qy 306 oPheSerPheThrGlnGlnAspMetProLysLeuArgArgGlnIle 321
      ||| |||||:::|||| :::: |||
Db 3249 TCTTGAATTATTAACAAGTAGATTTGAAGTCACTCATTCGCGCGACAATTA 3294
      ||| |||||:::|||| :::: |||

RESULT 13
US-09-134-001C-1230
; Sequence 1230, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1230
; LENGTH: 2184
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1230

Alignment Scores:
Pred. No.: 1.47 Length: 2184
Score: 86.00 Matches: 49
Percent Similarity: 36.26% Conservative: 46
Best Local Similarity: 18.70% Mismatches: 93
Query Match: 4.85% Indels: 74
DB: 4 Gaps: 9

US-09-848-852A-3 (1-332) x US-09-134-001C-1230 (1-2184)

Qy 69 GluGluHisValThr---CysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGly 87
      |||||:::|||| |||||:::|||| :::: |||
Db 58 GAGGAATACATCAAGCTGTTGCTAGAGAGTATTAAAGCAACACGCTACACAGATTTAAT 117
      |||||:::|||| |||||:::|||| :::: |||
Qy 88 SerLeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGln 107
      :::: |||

```

```
Db 118 TTGATTATT-----GTGAAT 132
Qy 108 GluPheSerThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArg 127
Db 133 GATGGGTGCACAGACCAAGTAAATAATGATGAGCTATTAAAGACTATGATRAA 192
Qy 128 MetProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThr 147
Db 193 -----AACATTCGATTTTATCGATT 213
Qy 148 MetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMet 167
Db 214 GATGAAATAGTGGTCACATCGCGCTAATATTCGATTAGAGAAGTAGAACACCT 273
Qy 168 TyrGlyAspLeuValMetAspThrValProGluLysValHisPheAsnSerPhePhe 187
Db 274 TAC-----TTATGCTTTTAGATCGACATGATGAATTAGCTTCTTATGCAATCACGTTTAC 330
Qy 188 TyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTyrTrpLysAsnValAsp 207
Db 331 TTAGAAAAGTTTAAT----- 345
Qy 208 IlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHisTrpSerLeuIle 227
Db 346 -----AATACCTGATGGTTTAACTCTCAATACATCTTTTACTACACAACGTCGCA 399
Qy 228 SerValAspValArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArg 247
Db 400 TTTGTTGATTAGTCGAGTAGGTTGAATCTTAATGCCAAGAGAAATATTATCA 459
Qy 248 ArgCysProLysHisIleAlaLysTyrLeu-----GlnAlaGluAlaValLysLysAsp 265
Db 460 TTTTAAAGAAAACAGCTGCTGTGAACATTAATTTTAGAACAGCTATTGTCAGGCGCAT 519
Qy 266 ArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAsn 285
Db 520 CATATTCCGATTTAATGAAATTTAAATACATATGATGGATTGG----- 561
Qy 286 AsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHisLeu----- 301
Db 562 -----TCATTGCTCTGAAATATGAAGTATGTAATAAGTTTGTGA 603
Qy 302 ---AlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgGln 320
Db 604 CGTATTTTAACTTCCCATTTTATTC-----CGTGGGGA 639
Qy 321 IleTyr 322
Db 640 GTCTAT 645
RESULT 14
US-08-531-927B-8
; Sequence 8, Application US/08531927B
; Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491
; NUMBER OF INVENTION: Disease Gene and Uses Thereof
; CORRESPONDENCE ADDRESSES: 23
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/531.927B
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP H6-251600
; FILING DATE: 21-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: ATH95-01A
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 925 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-531-927B-8
Alignment Scores:
Pred. No.: 0.672 Length: 925
Score: 84.00 Matches: 45
Percent Similarity: 38.98% Conservative: 24
Best Local Similarity: 25.42% Mismatches: 53
Query Match: 4.74% Indels: 55
DB: 2 Gaps: 10
US-09-848-852A-3 (1-332) x US-08-531-927B-8 (1-925)
Qy 60 LysAlaGlyGlnHisSerProLeuArgGluGluHisValThr-CysValGlnSerIleLe 79
Db 49 AAACATGGAGTCCATCTCCACGAGAAAACAAGAGGCTACCTTGTGTCACACATGGCT 108
Qy 79 uAspGluPheLeuGln---ThrTyrGlySerLeuIleProLeuSerThrAspGluValVa 98
Db 109 GAATAACTTATTCGAAGGAGATATTTAGCCCTGTGGATTTATCTCA-----ATTGC 162
Qy 98 LclLysLeuGluAsp-----ProSerArgLys-----IlePheGlnGlnGln 108
Db 163 ACATCAGCTGGATGAGGAGGAGGATGAGATGCGAAGAGGAGGAGTACTAGGAAGA 222
Qy 108 upheserthr-----ProSerArgLys----- 115
Db 223 TTATCGCACGGTGTGCACAGACGCTTCTGGAATATGGATGACAGTGGTTTTTCTCTAT 282
Qy 116 -----GlyLeuValLeuGlnLeuIleGlnSer-- 124
Db 283 TCAGGTTATACCAATGCCTTCAAAAGTTTGGGTTTAGAACTAATCCTGTTCAACAGTCC 342
Qy 125 ---TyrGlnArgMet-----ProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 343 AGAGTATCAGAGGCTTAGGATCGATCCCTATAAATGAA-----AGATCATTTATATGCAA 396
Qy 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160
Db 397 TTATAAGGAACACTGTTTACAGTTAGAAAATTA-----GGAAAAACAGTGGTT 444
Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
Db 445 TAACTTGATTTCTCTTCAGCGGGTCCAGAAATTAATATCAGATACA----- 490
Qy 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyr 196
Db 491 -----TATCTTGCACTTTTCTTGCTCAATTAACAACAGGAAGGTAT 532
RESULT 15
US-08-902-632-1
; Sequence 1, Application US/08902632
; Patent No. 6008025
; GENERAL INFORMATION:
```

APPLICANT: KOMATSUBARA, Hideyuki  
 APPLICANT: KITABAYASHI, Masao  
 APPLICANT: KAWAMURA, Hideki  
 APPLICANT: KAWAKAMI, Bunsei  
 APPLICANT: KAWAMURA, Yoshihisa  
 APPLICANT: TAKAGI, Masahiro  
 APPLICANT: IMANAKA, Tadayuki  
 TITLE OF INVENTION: Modified Thermostable DNA Polymerase,  
 and DNA Polymerase Composition for Nucleic Acid  
 TITLE OF INVENTION: Amplification  
 NUMBER OF SEQUENCES: 23  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Kenyon & Kenyon  
 STREET: 1 Broadway  
 CITY: New York  
 STATE: NY  
 COUNTRY: US  
 ZIP: 10004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.50 inch floppy disk, 1.44 Mb storage  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: WordPerfect 6.1 Windows  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/902.632  
 FILING DATE: Concurrent Herewith  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 198911/96  
 FILING DATE: 29-JUL-1996  
 APPLICATION NUMBER: JP 200446/96  
 APPLICATION NUMBER: 30-JUL-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: GREASON, Edward W.  
 REGISTRATION NUMBER: 18,918  
 REFERENCE/DOCKET NUMBER: 2418/7  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-425-7200  
 TELEFAX: 212-425-5288  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5342 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: genomic DNA  
 ORIGINAL SOURCE:  
 ORGANISM: hyperthermophilic archaeon  
 STRAIN: KOD1  
 US-08-902-632-1

Qy	61	AlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAsp	80
Db	3792	-----GAGTTCGTCTTCAGTCCCGCAAGGGGTTGCGGT	3837
Qy	81	GlupHeLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluValValGluLys	100
Db	3828	:::        :::      ::::::	3887
Qy	101	LeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGln	120
Db	3888	:::    :::    :::    :::    :::	3917
Qy	121	LeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGly-	136
Db	3918	CTGCTCCTCTTTGAACCTCGGTGGGGGTCTCTGCTGTAAACATTGGGCACGACGACGGC	3977
Qy	137	--PheArgVal-	139
Db	3978	GTTTACAGGGTCTATATAACGAGGAGCTCCCGTTTCGTAAAGCTGGACAAGAAAAAGAAC	4037
Qy	140	AlaTyrlYsArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrglyGln-	157
Db	4038	:::    :::    :::      :::	4097
Qy	157	-----	157
Db	4098	CAGAAAAACGTCTAGTCCTCAGACCTTCAGGAAGATGGTTCAGGACGGGAAGACTCGATCCC	4157
Qy	158	-----AsnTrpLeuAsnAspGlnValMetAsnMetTyrglyAspLeu	171
Db	4158	GAAGAAGCCCGAGAGGCTCTCCTGGCTC-----ATTGAGGGGAGCGTA	4199
Qy	172	ValMetAspThrValProGlu-	178
Db	4200	GTGCTGACCGCGCTTGAGTCGCTGATGTGAAGACTACGATGGTTATGCTATGACCTG	4259
Qy	179	-----LysValHisPhePheAsnSer	185
Db	4260	AGCGTCGAGGACAAACAGAACTCTCCGTTGGCTTGGTTGGTCTATGCTCACAACAGC	4319
Qy	186	Phe-----PheTyrlAspLysLeuArgThrLysGlyTyrlAspGlyValLysArgTrpThr	203
Db	4320	TACTACGGTTTACTACGGCTATCAAGAGCGCGCTCGGTAC-----TCCAGGAGTGTCGA	4373
Qy	204	LysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIle---HisLeuGluVal	222
Db	4374	GAGAGCGTAACGCCTGGGAGGAGGTACATAACGATGACCATCAAGGAGATAGAGAA	4433
Qy	223	HisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrlPheAspSerGln	242
Db	4434	AAGTACGGGCTTT-----AAGGTAATCTACAGCGACACCGCAC	4469
Qy	243	-----ArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrlLeuGlnAla	259
Db	4470	GGATTTTTGCCACAATA-----CCTGGACCGCATGCT	4502
Qy	260	GluAlaValLysLysAspArgLeuAspPhe-----	269
Db	4503	GAACCGTCAAAAAGGAGGCTATGGAGTTCTCTCAACTATATCAACGCCAAACTTCCGGGC	4562
Qy	270	-----HisGlnClyTrpLysGlyTyrlPheLysMetAsnValAlaArgGlnAsnAsp	287
Db	4563	GCCTTTGAGCTCGAGTACGAGGGCTTCTACAAACGC-----	4598
Qy	288	SerAspCysGlyAlaPheValLeuGlnTyrlCysLysHisLeuAlaLeuSerGlnProPhe	307
Db	4599	-----GGCTTCTCTCGTCAGGAAG-----AAGTATGGGTGATAGCAGGAAGGC	4646
Qy	308	SerPheThrGlnGlnAspMetProLysLeuArgArg-----GlnIleTyrlLysGlu	324
Db	4647	AAGATAACACCGCGGACTCTGAGATTGTGAGCGCTGACTGGAGCGAGATAGCAAGAG	4706



Search completed: December 16, 2002, 20:17:51  
Job time : 115 secs

1

1

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 19:34:35 ; Search time 771 Seconds  
(without alignments)  
88.726 Million cell updates/sec

Title: US-09-848-852A-3  
Perfect score: 1772  
Sequence: 1 MYSQRFWGTIWARAGHA.....DMPKLRROIYKELCHCKLTV 332

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rviris.\*  
16: sp\_bacteriap.\*  
17: sp\_archheap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1667	94.1	371	4 Q9Y3W9	Q9Y3W9 homo sapien
2	1667	94.1	574	4 Q9H4L4	Q9H4L4 homo sapien
3	1667	94.1	574	4 Q96PS4	Q96PS4 homo sapien
4	1637	92.4	568	11 Q9EP97	Q9EP97 mus musculus
5	789	44.5	537	4 Q96H10	Q96H10 homo sapien
6	788.5	44.5	755	6 Q9WP32	Q9WP32 macaca fasc
7	784	44.2	446	4 Q96SA5	Q96SA5 homo sapien
8	490	27.7	191	4 Q9NWF3	Q9NWF3 homo sapien
9	456	25.7	643	4 Q9P003	Q9P003 homo sapien
10	393	22.2	412	4 Q9P2L5	Q9P2L5 homo sapien
11	393	22.2	589	4 Q96SR2	Q96SR2 homo sapien
12	393	22.2	590	4 Q9HC62	Q9HC62 homo sapien
13	387.5	21.9	357	11 Q9CUM5	Q9CUM5 mus musculus
14	385.5	21.8	541	11 Q91XK6	Q91XK6 mus musculus
15	385.5	21.8	588	11 Q9D4Z0	Q9D4Z0 mus musculus
16	380	21.4	588	11 Q9EQE1	Q9EQE1 rattus norv

17	311	17.6	468	5 Q9VP10	Q9VP10 drosophila
18	307	17.3	1513	5 Q9VWK5	Q9VWK5 drosophila
19	282.5	15.9	497	10 Q9FP66	Q9FP66 oryza sativ
20	266.5	15.0	155	11 Q9WV12	Q9WV12 mus musculus
21	260.5	14.7	478	10 Q9M908	Q9M908 arabidopsis
22	260.5	14.7	489	10 Q94F30	Q94F30 arabidopsis
23	253.5	14.3	100	6 Q95KG4	Q95KG4 macaca fasc
24	253.5	14.3	244	5 Q8SVK1	Q8SVK1 encephalito
25	252	14.2	157	5 O01650	O01650 drosophila
26	240.5	13.6	233	10 Q65278	Q65278 arabidopsis
27	201.5	11.4	571	10 Q8RWNO	Q8RWNO arabidopsis
28	201.5	11.4	582	10 Q9X1J4	Q9X1J4 arabidopsis
29	197.5	11.1	424	10 Q23439	Q23439 arabidopsis
30	193	10.9	547	10 Q80745	Q80745 arabidopsis
31	180	10.2	506	10 Q8S2J1	Q8S2J1 oryza sativ
32	173.5	9.8	1017	4 Q96PS5	Q96PS5 homo sapien
33	169.5	9.6	1148	5 Q9VRY4	Q9VRY4 drosophila
34	163	9.2	710	10 Q81879	Q81879 arabidopsis
35	156	8.8	1240	3 Q9P6U5	Q9P6U5 neuropeptid
36	154	8.7	560	5 Q9VYJ5	Q9VYJ5 drosophila
37	143.5	8.1	1011	10 Q8S1I4	Q8S1I4 oryza sativ
38	143	8.1	1105	4 Q8TBY4	Q8TBY4 homo sapien
39	137.5	7.8	997	10 Q8RZJ5	Q8RZJ5 oryza sativ
40	137.5	7.8	1007	10 Q94LG5	Q94LG5 oryza sativ
41	136.5	7.7	1046	10 Q8W0H2	Q8W0H2 oryza sativ
42	136	7.7	893	5 Q23238	Q23238 caenorhabdi
43	135.5	7.6	1053	10 Q94E62	Q94E62 oryza sativ
44	135	7.6	226	10 Q9LSS7	Q9LSS7 arabidopsis
45	135	7.6	1011	10 Q8W5N9	Q8W5N9 oryza sativ

ALIGNMENTS

RESULT 1

Q9Y3W9 PRELIMINARY; PRT; 371 AA.  
AC Q9Y3W9;  
DT 01-NOV-1999 (TREMREL. 12, Created)  
DT 01-NOV-1999 (TREMREL. 12, Last sequence update)  
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)  
DE Hypothetical 42.2 kDa protein.  
GN DKFZP586K0919.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=UTERUS;  
RX MEDLINE=21154917; PubMed=11230166;  
RA Wiemann S., Weil B., Wellenreuther R., Cassenhuber J., Glassl S.,  
RA Ansgore W., Boecher M., Bloecher H., Bauersachs S., Blum H.,  
RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,  
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,  
RA Wambutt R., Korn B., Klein M., Poustka A.;  
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and  
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";  
RL Genome Res. 11:422-435(2001).  
DR EMBL; AL050283; CAB43384.1; -.  
DR HSSP; Q02724; 1EUV.  
DR MEROPS; C48.003; -.  
DR InterPro; IPR003653; SUMO\_protease.  
DR Pfam; PF02902; Peptidase\_C48; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 371 AA; 42234 MW; EEE87C5885CA5082 CRC64;  
  
Query Match 94.1%; Score 1667; DB 4; Length 371;  
Best Local Similarity 100.0%; Pred. No. 1.7e-136; Indels 0;  
Matches 314; Conservative 0; Mismatches 0; Gaps 0;  
  
QY 19 LAPPDASILISVCSIGDHVAQELFGQSDLGMAEAEAEPEKAGQHSPLREHVTVCQSI 78  
|||||

```
Db 58 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 117
QY 79 LDEFLOTYGLSLPLSTDEVVEKLEDFIQQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 138
Db 118 LDEFLOTYGLSLPLSTDEVVEKLEDFIQQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 177
QY 139 VAYKRHLVTMDLGLTYGQNLNDQVNMVYGDVMTVPKLVHFFNSFFYDKLRTKGYDG 198
Db 178 VAYKRHLVTMDLGLTYGQNLNDQVNMVYGDVMTVPKLVHFFNSFFYDKLRTKGYDG 237
QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 258
Db 238 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 297
QY 259 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLQYCKHLALSOPFSFTQODMPKLR 318
Db 298 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLQYCKHLALSOPFSFTQODMPKLR 357
QY 319 RQIYKELCHCKLTV 332
Db 358 RQIYKELCHCKLTV 371

RESULT 2
Q9H4L4
ID Q9H4L4 PRELIMINARY; PRT; 574 AA.
AC Q9H4L4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sentrin/SUMO-specific protease.
GN SENP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA Choi S.J., Jeon Y.J., Kim K.I., Nishimori S., Suzuki T., Uchida S.,
RA Shimbara N., Tanaka K., Chung C.H.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199459; AAL25652.1; -
DR MEROPS; C48.003; -
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Protease.
SQ SEQUENCE 574 AA; 64856 MW; 830FD39C4D02C0EB CRC64;

Query Match 94.1%; Score 1667; DB 4; Length 574;
Best Local Similarity 100.0%; Pred. No. 3.2e-136;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 78
Db 261 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 320
QY 79 LDEFLOTYGLSLPLSTDEVVEKLEDFIQQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 138
Db 321 LDEFLOTYGLSLPLSTDEVVEKLEDFIQQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 380
QY 139 VAYKRHLVTMDLGLTYGQNLNDQVNMVYGDVMTVPKLVHFFNSFFYDKLRTKGYDG 198
Db 381 VAYKRHLVTMDLGLTYGQNLNDQVNMVYGDVMTVPKLVHFFNSFFYDKLRTKGYDG 440
QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 258
Db 441 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 500
QY 259 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLQYCKHLALSOPFSFTQODMPKLR 318
Db 501 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLQYCKHLALSOPFSFTQODMPKLR 560

Query Match 94.1%; Score 1667; DB 4; Length 574;
Best Local Similarity 100.0%; Pred. No. 3.2e-136;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 78
Db 261 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 320
QY 79 LDEFLOTYGLSLPLSTDEVVEKLEDFIQQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 138
Db 321 LDEFLOTYGLSLPLSTDEVVEKLEDFIQQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 380
QY 139 VAYKRHLVTMDLGLTYGQNLNDQVNMVYGDVMTVPKLVHFFNSFFYDKLRTKGYDG 198
Db 381 VAYKRHLVTMDLGLTYGQNLNDQVNMVYGDVMTVPKLVHFFNSFFYDKLRTKGYDG 440
QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 258
Db 441 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 500
QY 259 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLQYCKHLALSOPFSFTQODMPKLR 318
Db 501 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLQYCKHLALSOPFSFTQODMPKLR 560
```

```
QY 319 RQIYKELCHCKLTV 332
Db 561 RQIYKELCHCKLTV 574

RESULT 3
Q96PS4
ID Q96PS4 PRELIMINARY; PRT; 574 AA.
AC Q96PS4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SUMO-1 specific protease 3.
GN SSP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA Choi S.J., Jeon Y.J., Kim K.I., Nishimori S., Suzuki T., Uchida S.,
RA Shimbara N., Tanaka K., Chung C.H.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199459; AAL25652.1; -
DR MEROPS; C48.003; -
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Protease.
SQ SEQUENCE 574 AA; 65009 MW; E495137EE7500741 CRC64;

Query Match 94.1%; Score 1667; DB 4; Length 574;
Best Local Similarity 100.0%; Pred. No. 3.2e-136;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 78
Db 261 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 320
QY 79 LDEFLOTYGLSLPLSTDEVVEKLEDFIQQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 138
Db 321 LDEFLOTYGLSLPLSTDEVVEKLEDFIQQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 380
QY 139 VAYKRHLVTMDLGLTYGQNLNDQVNMVYGDVMTVPKLVHFFNSFFYDKLRTKGYDG 198
Db 381 VAYKRHLVTMDLGLTYGQNLNDQVNMVYGDVMTVPKLVHFFNSFFYDKLRTKGYDG 440
QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 258
Db 441 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 500
QY 259 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLQYCKHLALSOPFSFTQODMPKLR 318
Db 501 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLQYCKHLALSOPFSFTQODMPKLR 560

Query Match 94.1%; Score 1667; DB 4; Length 574;
Best Local Similarity 100.0%; Pred. No. 3.2e-136;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 RQIYKELCHCKLTV 332
Db 561 RQIYKELCHCKLTV 574

RESULT 4
Q9EP97
ID Q9EP97 PRELIMINARY; PRT; 568 AA.
AC Q9EP97;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sentrin/SUMO-specific protease (SMT3 isopeptidase 1).
GN SMT3IP1 OR SMT3IP.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C57BL/6J;
RX  MEDLINE=20267842; PubMed=10806345;
RA  Yeh E.T., Gong L., Kamitani T.;
RT  "Ubiquitin-like proteins: new wines in new bottles.";
RL  Gene 248:1-14(2000).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  PubMed=11029585;
RA  Nishida T., Tanaka H., Yasuda H.;
RT  "A novel mammalian Smt3-specific isopeptidase 1 (SWT3IP1) localized in
the nucleolus at interphase.";
RL  Eur. J. Biochem. 267:6423-6427(2000).
DR  EMBL; AY008764; AAG33253.1; -
DR  EMBL; AF194031; AAG28418.1; -
DR  HSSP; Q02724; IEUV.
DR  MEROPS; C48.003; -.
DR  MGD; MGI:2158736; Smt3ip1.
DR  InterPro; IPR003653; SUMO_protease.
DR  Pfam; PF02902; Peptidase_C48; 1.
KW  Protease.
SQ  SEQUENCE 568 AA; 64403 MW; 655F1FABIAB62EA8 CRC64;

Query Match 92.4%; Score 1637; DB 11; Length 568;
Best Local Similarity 98.1%; Pred. No. 1.3e-133;
Matches 308; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY  19 LAPPDASILISNVCISGIDHVAQELFGQSDLGMAEAEAPRGKAGQHSPLREEHVTCVQSI 78
DB  255 LAPPDASILISNVCISGIDHVAQELFGQSDLGIAEADRTGKAGQHSPLREEHVTCVQSI 314
QY  79 LDEFLOTQYGLSLPLSTDEVEKLEDFQEEFSTPSRKGLVLQLIQSYORMPGNMVRGR 138
DB  315 LDEFLOTQYGLSLPLSTDEVEKLEDFQEEFSTPSRKGLVLQLIQSYORMPGNMVRGR 374
QY  139 VAYKRHLVMTDGLTLYGQNLNDQVMNMYGDLVMDTVPEKVFHFFNSFYDKLRTKGYDG 198
DB  375 VSYKRHLVMTDGLTLYGQNLNDQVMNMYGDLVMDTVPEKVFHFFNSFYDKLRTKGYDG 434
QY  199 YKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTTYFDQSRTLNRCPKHIKYLQ 258
DB  435 YKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTTYFDQSRTLNRCPKHIKYLQ 494
QY  259 AEAVKKDLRDFHQGWKGYFKMNVARQNDSDCGAFVLYQYCKHLALSQPFSFTQDDMPKLR 318
DB  495 AEAVKKDLRDFHQGWKGYFKMNVARQNDSDCGAFVLYQYCKHLALSQPFSFTQDDMPKLR 554
QY  319 ROIYKELCHCKLTIV 332
DB  555 ROIYKELCHCKLTIV 568

RESULT 5
Q96H10 PRELIMINARY; PRT; 537 AA.
AC Q96H10;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Unknown (Protein for IMAGE:3448367) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008589; AA008589.1; -
DR MEROPS; C48.008; -.
DR InterPro; IPR003653; SUMO_protease.

```

```

DR Pfam; PF02902; Peptidase_C48; 1.
FT NON_TER 1
SQ SEQUENCE 537 AA; 61769 MW; 18F609C41D4B7DF9 CRC64;

Query Match 44.5%; Score 789; DB 4; Length 537;
Best Local Similarity 54.0%; Pred. No. 5.1e-60;
Matches 148; Conservative 54; Mismatches 66; Indels 6; Gaps 3;

QY  58 GEKAGQHSPLREHVT-CVQSTLDEFLOTQYGLSLPLSTDEVEKLEDFQEEFSTPSRK 116
DB  267 GNSKASQSPVDDEQUSVCLSGFLDEVMKYGSLVPLSEKEVLGRKLDVFNEDFS--NRKP 324
QY  117 LVLQLIQSYQRMFGNAMVRGFRVAYKRHLVMTDGLTLYGQNLNDQVMNMYGDLVMDTV 176
DB  325 FINREITNVRARHKC---NFRIFYNKHMLDMDLATLDGQNLNDQVINMYGELIMDAV 381
QY  177 PEKVHFFNSFYDKLRTKGYDGKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTI 236
DB  382 PDKVHFFNSFYHRLVTKGYNGVKRWTKVDLFRKSLLLIPILHLEVHWSLITVTLNRII 441
QY  237 TYFDSQRTLNRCPKHIKYLQAEAVKKDLRDFHQGWKGYFKMNVARQNDSDCGAFVLYQ 296
DB  442 SFYDSQGIHFKFCVENIRKYLTEAREKNRPEFLQGWQVAVTKCIPQQKNDSDCGVFLVQ 501
QY  297 YCKHLALSQPFSFTQDDMPKLRRIYKELCHCKL 330
DB  502 YCKLALQPPFQSOEDMPRVKRIYKELCECRL 535

RESULT 6
Q8WP32 PRELIMINARY; PRT; 755 AA.
AC Q8WP32;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical 86.3 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
libraries.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB074445; BAB72076.1; -
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Hypothetical protein.
SQ SEQUENCE 755 AA; 86290 MW; 40EC773CA29B8CEA CRC64;

Query Match 44.5%; Score 788.5; DB 6; Length 755;
Best Local Similarity 53.5%; Pred. No. 9.2e-60;
Matches 152; Conservative 53; Mismatches 72; Indels 7; Gaps 4;

QY  49 GMAEEAERPGKAGQH-SPLREHVT-CVQSTLDEFLOTQYGLSLPLSTDEVEKLEDFQ 106
DB  475 GLKLENQVGGKDSQKASFPVDEQLSVCLSGFLDEVMKYGSLVPLSEKEVLGRKLDVFN 534
QY  107 QEFSTPSRKGLVLQLIQSYQRMFGNAMVRGFRVAYKRHLVMTDGLTLYGQNLNDQVMN 166
DB  535 EDFS--NRKPFINREITNVRARHKC---NFRIFYNKHMLDMDLATLDGQNLNDQVIN 589
QY  167 MYGDLVMDTVPEKVHFFNSFYDKLRTKGYDGKRWTKNVDIFNKELLIPILHLEVHWSL 226
DB  590 MYGELIMDAVPDKVHFFNSFYHRLVTKGYNGVKRWTKVDLFRKSLLLIPILHLEVHWSL 649
QY  227 ISVDVRRRTIYFDQSRTLNRCPKHIKYLQAEAVKKDLRDFHQGWKGYFKMNVARQNN 286

```

```

Db 650 ITVTLNRIISFYDSQIHFKFCVENIRKYLLEAREKNRPEFLQHQWQTAVTKCIPOOKN 709
QY 287 DSDCGAVLOYCKHALSQSFSTQDMPKLRQIYKELCHCKL 330
Db 710 DSDCGVEVLOYCKLALEQPFQSFQSDMPRVRRIYKELCECL 753

RESULT 7
Q96SA5 PRELIMINARY; PRT; 446 AA.
ID Q96SA5
AC Q96SA5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sumo/sentrin-specific protease.
GN FKSG45.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang Y.-G., Li T.;
RT "Identification of FKSG45, a novel gene located on human chromosome
RT 3."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL: AF335474; AK69630.1;
DR MEROPS; C48.008;
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Protease.
SQ SEQUENCE 446 AA; 50902 MW; 3810DCC23655D646 CRC64;

Query Match 44.2%; Score 784; DB 4; Length 446;
Best Local Similarity 53.6%; Pred. No. 1.1e-59;
Matches 147; Conservative 54; Mismatches 67; Indels 6; Gaps 3;

QY 58 GERAGQHSPLREHVT-CVQSIIDFLQTYGSLIPSTDEVEKLEDFQOEFSTPSRK 116
Db 176 GKNSQKASPVDDQSLVCLSGFLDEWKKYKGLSLVSEKEVLGRKDKVFNEDFC--NRKP 233
QY 117 LVLIQLQSYQRMFGNMGVGRVAYKRHLVLTMDLGLTYGQNNLNDQVMNMGDLVMDTV 176
Db 234 FINREITNRYARKQC--NFRIFYKNHMLDMDLATLDGQNLNDQVINMGELIMDAV 290
QY 177 PEKVHFNFSFYDKLRTKGDGVKRTKNVDIENKELLLIPIHLEVHWSLISVDVRRRTI 236
Db 291 PDKVHFNFSFRLQVTKGYNGVGRWTKKYDLEFKKSLLLIPIHLEVHWSLITVLSNRII 350
QY 237 TYFDSORTLNRCPKHTAKYLOAEAVKDRDLDFHOGKMGYFKMNVARQNNDSDCGAFVLQ 296
Db 351 SFYDSQGIHFKFCVENIRKYLLEAREKNRPEFLQHQWQTAVTKCIPOOKNDSDCGAFVLQ 410
QY YCKHLALSQPSFTQDMPKLRQIYKELCHCKL 330
Db 411 YCKCLALEQPFQSFQSDMPRVRRIYKELCECL 444

RESULT 8
Q9NWF3 PRELIMINARY; PRT; 191 AA.
ID Q9NWF3
AC Q9NWF3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Hypothetical 20.5 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RC TISSUE=EMBRYO;
RA Isoai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL: AK00923; BAA91428.1;
SQ SEQUENCE 191 AA; 20457 MW; B9F2B6F629F31612 CRC64;

Query Match 27.7%; Score 490; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.1e-34;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILISNVCISGDHVAQELFOGSDLGMAEAEERPGEKAGQHSPLREHVTQVSI 78
Db 96 LAPPDASILISNVCISGDHVAQELFOGSDLGMAEAEERPGEKAGQHSPLREHVTQVSI 155
QY 79 LDEFLOTYGSLLPLSTDEVVEKLEDFQOEFSTPSR 114
Db 156 LDEFLOTYGSLLPLSTDEVVEKLEDFQOEFSTPSR 191

RESULT 9
Q9P0U3 PRELIMINARY; PRT; 643 AA.
ID Q9P0U3
AC Q9P0U3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sentrin/SUMO-specific protease.
GN SENP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20119292; PubMed=10652325;
RA Gong L., Millas S., Maul G.G., Yeh E.T.H.;
RT "Differential Regulation of Sentrinized Proteins by a Novel Sentrin-
RT specific Protease.";
RL J. Biol. Chem. 275:3355-3359(2000).
DR ENBL: AF149770; AAF31171.1;
DR HSSP; Q02724; LEUV.
DR MEROPS; C48.002;
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Protease.
SQ SEQUENCE 643 AA; 73351 MW; 1556864F1BDEE337 CRC64;

Query Match 25.7%; Score 456; DB 4; Length 643;
Best Local Similarity 34.0%; Pred. No. 5.5e-31;
Matches 112; Conservative 60; Mismatches 111; Indels 46; Gaps 11;

QY 24 ASLISNVCISGDHVAQE----LFQSGSDLGMAEAEERPGEKAGQHS-----PLR 68
Db 338 AELWIKELTSVYDSRARELRQIEQKALQALQNLQQLQER--EHSVHDSVELHLRVPLE 395
QY 69 EE-HVTCVQSILDEFLOTYGSLLPLSTDEVVEKLEDFQOEFSTPSRKGLVLQLQSYOR 127
Db 396 KEIPVTIVQVE-----TQKKGKHLTDESEDFEPEITEM-EKEIKNVFRNG----- 438
QY 128 MFGNAMVGRFVAYKRHLVLTMDLGLTYGQNNLNDQVMNMGDLVMDTVPEK----VHFF 183
Db 439 NQDEVLSAEFL-----TITRKDIQTLNHLNWLNDIINFYNNMLMERKSGKGLPSVHAF 493
QY 184 NSFFYDKLTKTGVDGVKRWTKNVDIENKELLLIPIHLEVHWSLISVDVRRRTTYDSOR 243
Db 494 NTFFFTKLKTAGYQAVKRWTKVDVFSVDILLVPIHLGVHCLAVVDYFRKKNITYDSMG 553
```



[illegible]



```
Db 319 -EFTDMEKEI-----SNALGHGPPDEILSSAFKLRI-TRGDIQTLKYNHWLN 364
QY 162 DOVMNMYGDLVMDTVPEK----VHFFNSFFYDKLRTKGYDGVKRWTKNVDIENKELLIP 217
Db 412 DEVINFYNLLVERSKKQGYPALHAFSTFFYPKLKSGGYQAVKRWTKGVNLFQELVLP 471
QY 218 IHLEVHWSLISVDVRRRTITYPDSQRTLNRRCPKHIAYQLQAEAVKKDRLDHF-OGWKY 276
Db 472 IHRKVHWSLVVMDLRKCKLYLDSMGQKGRHICEILLQYLQDESCKTKRNTDNLLEWTHY 531
QY 277 FKM--NVARQNNDSDCGAFVLYQYCKHLALSOPFSFTQODMPKLRROIYKELCHCKL 330
Db 532 SMKPEIPOOLNGSDCGMFTCKYADYISRDKITTFTHQOMPLFRKKMWWEILHQOL 587

Search completed: December 16, 2002, 20:11:15
Job time : 773 secs

RESULT 15
Q9D4Z0
ID Q9D4Z0 PRELIMINARY; PRT; 588 AA.
AC Q9D4Z0;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE 4930538C18Rik protein.
GN 4930538C18Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., StaUBL F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Rodriguez I., Sakamoto N.,
RA Nordone P., King B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015987; BAB30067.1; -.
DR HSSP; Q02724; 1EUV.
DR MEROPS; C48.007; -.
DR MGD; MGI:1923076; 4930538C18Rik.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
SQ SEQUENCE 588 AA; 67579 MW; 09B56796CA194847 CRC64;

Query Match
Best Local Similarity 21.8%; Score 385.5; DB 11; Length 588;
Matches 93; Conservative 62; Mismatches 106; Indels 35; Gaps 9;

QY 55 ERPGKAGQHSPLREHVTCT--VQSILDEFLLQYCSLIPLS-----TDEVVKEKIDIFQ 106
Db 307 EKEGTRGHQMEFDLSEVSARLGLSGSNGLLRRKISLVLEIKEKNFPSPKEDRPTEDLF- 365
QY 107 QEFSTPSRKGLVLQLIQSVQRMGNMVRG-----FRVAYKRHLVTMDDLGTLGYQNWLN 161
Db 366 -EFTDMEKEI-----SNALGHGPPDEILSSAFKLRI-TRGDIQTLKYNHWLN 411
```

```
QY 162 DOVMNMYGDLVMDTVPEK----VHFFNSFFYDKLRTKGYDGVKRWTKNVDIENKELLIP 217
Db 412 DEVINFYNLLVERSKKQGYPALHAFSTFFYPKLKSGGYQAVKRWTKGVNLFQELVLP 471
QY 218 IHLEVHWSLISVDVRRRTITYPDSQRTLNRRCPKHIAYQLQAEAVKKDRLDHF-OGWKY 276
Db 472 IHRKVHWSLVVMDLRKCKLYLDSMGQKGRHICEILLQYLQDESCKTKRNTDNLLEWTHY 531
QY 277 FKM--NVARQNNDSDCGAFVLYQYCKHLALSOPFSFTQODMPKLRROIYKELCHCKL 330
Db 532 SMKPEIPOOLNGSDCGMFTCKYADYISRDKITTFTHQOMPLFRKKMWWEILHQOL 587

Search completed: December 16, 2002, 20:11:15
Job time : 773 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 19:26:20 ; Search time 167 Seconds  
(without alignments)  
33.125 Million cell updates/sec

Title: US-09-848-852A-3

Perfect score: 1772

Sequence: 1 MYSQRFWTGIWARRGAHLA.....DMPKLRRIQYKELCHCKLTIV 332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105981 seqs, 16662342 residues

Total number of hits satisfying chosen parameters: 105981

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1772	100.0	332	10	US-09-848-852A-3
2	222.5	12.6	621	10	US-09-856-247A-2
3	122.5	6.9	212	12	US-10-008-461-2
4	90.5	5.1	427	10	US-09-815-242-13640
5	89.5	5.1	427	10	US-09-815-242-13443
6	86	4.9	824	10	US-09-866-582-34
7	84	4.7	304	10	US-09-925-302-614
8	82.5	4.7	674	10	US-09-765-272-200
9	82	4.6	803	10	US-09-801-368-394
10	80	4.5	439	9	US-09-890-813-16
11	80	4.5	865	10	US-09-815-242-11364
12	79	4.5	638	10	US-09-815-242-5222
13	79	4.5	642	10	US-09-815-242-12143
14	78.5	4.4	782	9	US-10-041-007-16
15	78.5	4.4	782	10	US-09-887-586A-46
16	78.5	4.4	782	10	US-09-903-012-46
17	78.5	4.4	817	9	US-10-041-007-15
18	78.5	4.4	820	10	US-09-866-582-35
19	78.5	4.4	1503	9	US-10-007-706-1

Sequence 27, Appl  
Sequence 4912, Ap  
Sequence 10686, A  
Sequence 138, App  
Sequence 11557, A  
Sequence 24, Appl  
Sequence 24, Appl  
Sequence 5126, Ap  
Sequence 11522, A  
Sequence 142, App  
Sequence 2, Appl  
Sequence 9, Appl  
Sequence 50, Appl  
Sequence 7, Appl  
Sequence 52, Appl  
Sequence 52, Appl  
Sequence 22, Appl  
Sequence 11036, A  
Sequence 617, App  
Sequence 12984, A  
Sequence 2, Appl  
Sequence 184, App  
Sequence 10596, A  
Sequence 11536, A

Sequence 10  
US-09-905-173-27  
US-09-815-242-4912  
US-09-815-242-10686  
US-09-870-759-128  
US-09-815-242-11557  
US-09-887-586A-24  
US-09-903-012-24  
US-09-815-242-5126  
US-09-815-242-11522  
US-09-745-763-142  
US-09-819-249-2  
US-09-946-239-9  
US-09-991-496-50  
US-09-874-923-50  
US-09-938-956-7  
US-09-887-586A-52  
US-09-903-012-52  
US-09-887-586A-22  
US-09-903-012-22  
US-09-815-242-11036  
US-09-925-297-617  
US-09-815-242-12984  
US-10-047-757-2  
US-09-881-752A-184  
US-09-815-242-10596  
US-09-815-242-11536

#### ALIGNMENTS

RESULT 1  
US-09-848-852A-3  
; Sequence 3, Application US/09848852A  
; Patent No. US20020106373A1  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; Tang, Y. Tom  
; Corley, Neil C.  
; Guegler, Karl J.  
; Yue, Henry  
; Patterson, Chandra  
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH EGF SIGNALING  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/848,852A  
; FILING DATE: 07-May-2001  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/069,725  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0515 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 332 amino acids  
; TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BEPINOT01
; CLONE: 2056178
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-848-852A-3

Query Match      100.0%; Score 1772; DB 10; Length 332;
Best Local Similarity 100.0%; Pred. No. 3.6e-167;
Matches 332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYSQRFWGTIWARGAHLAPPDASILISNVCSIGDHVAQELFGSDLGMAEAEAPGPK 60
DB 1 MYSQRFWGTIWARGAHLAPPDASILISNVCSIGDHVAQELFGSDLGMAEAEAPGPK 60
QY 61 AQHSPLEHVTVCQSILDEFLQYGSILPLSTDEVEVEKLEDFIQQEFSTPSRGLVLQ 120
DB 61 AQHSPLEHVTVCQSILDEFLQYGSILPLSTDEVEVEKLEDFIQQEFSTPSRGLVLQ 120
QY 121 LIQSYORMPGNAMVRGRVAYKRHLVTMDLGLTLYGQNLNDQVMNMGDLVMDTVPEKV 180
DB 121 LIQSYORMPGNAMVRGRVAYKRHLVTMDLGLTLYGQNLNDQVMNMGDLVMDTVPEKV 180
QY 181 HFNFSFFYDKLRTKGYDGVKRWTKNVDIFNKELLIPIHLEVHWSLISVDVRRRTITYFD 240
DB 181 HFNFSFFYDKLRTKGYDGVKRWTKNVDIFNKELLIPIHLEVHWSLISVDVRRRTITYFD 240
QY 241 SORTLNRRCPKHIAKYLOAEVKKORLDPHQGWKGYFKMNVARQNNDSCGAFVLYQCKH 300
DB 241 SORTLNRRCPKHIAKYLOAEVKKORLDPHQGWKGYFKMNVARQNNDSCGAFVLYQCKH 300
QY 301 LALSOPFSFTQDMPKLRRIQYKELCHCKLTV 332
DB 301 LALSOPFSFTQDMPKLRRIQYKELCHCKLTV 332

RESULT 2
US-09-856-247A-2
; Sequence 2, Application US/09856247A
; Patent No. US20020151028A1
; GENERAL INFORMATION:
; APPLICANT: Lima, Christopher
; APPLICANT: Mossesova, Elena
; TITLE OF INVENTION: Structure-Based Drug Design for Ulpl Protease Substrates
; FILE REFERENCE: 2650/1G681-US1
; CURRENT APPLICATION NUMBER: US/09/856, 247A
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 60/205, 336
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-856-247A-2

Query Match      12.6%; Score 222.5; DB 10; Length 621;
Best Local Similarity 25.2%; Pred. No. 6.5e-14;
Matches 77; Conservative 56; Mismatches 120; Indels 53; Gaps 12;

QY 36 DHVAQEL-FQGSDLGMAEAEAPGKAGQHSPLREHVTVCQSILDEFLQYGSILPLST 94
DB 355 DYLNKLFDRSILFEKDFKRYNEILNERKKIQEDLKKKKQLAKK-----KLVp--- 405
QY 95 DEWEVEKLEDFIQQEFSTPSRGLVLQLQSYORMPGNAMVRGRVAYKRHLVTMDLGLT 154
DB 406 -ELNEKDDQVQKALA--SRENTQMLNRDNIE-----ITVRDEKTL 443
QY 155 YQCNWLNDQVMNMGDLVMDTVPEKVHFNFSFFYDKLRTKGYDGVKRWTK-----NVDLFN 210
DB 444 APRRWLNDTIIEFFMKYTEKSTPNTV-AFNSFFYTNLSERGQYGVRRWKRKKTKQIDKLD 502
```

```
QY 211 KELLIPHL-EVHWSLISVDVRRRTITYFDS-----QRTLNRRCPKHIAKYLOAEAVKDD 265
DB 503 K--IETPINLQSHWALGIIDLKKTIGYVDSLNSNGPNAMSFALITDLQKYVMEESKHTI 560
QY 266 RLDFHQGWKGYFKMNVARQNNDSCGAFVLYQCKHIALSQPFSFTQDMPKLRRIQYKEL 325
DB 561 GDFG-----DLIHLDCPOOPNGYDCGIYVCMNTLYGSADAPLDYDKAIRMRRTI---- 611
QY 326 CHCKLT 331
DB 612 AHLILT 617

RESULT 3
US-10-008-461-2
; Sequence 2, Application US/10008461
; Patent No. US20020127692A1
; GENERAL INFORMATION:
; APPLICANT: Barbara Ink
; APPLICANT: Alan Lewis
; TITLE OF INVENTION: New Protein
; FILE REFERENCE: QGI034(p80209)
; CURRENT APPLICATION NUMBER: US/10/008, 461
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 0027905.9
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-008-461-2

Query Match      6.9%; Score 122.5; DB 12; Length 212;
Best Local Similarity 23.9%; Pred. No. 0.0001;
Matches 47; Conservative 32; Mismatches 75; Indels 43; Gaps 7;

QY 139 VAYKRHLVTMDLGLTLYGQNLNDQVMNMGDLVMDTVPEKVHFNFSFFYDKLR----- 192
DB 6 LSYMDSLRLRQSDVSLDDPPSWLNDHIIGFAFE-----YFANSQFHDCSDHVSFIS 55
QY 193 -----TKGYDGVKRWTKNVDIFNKELLIPHL-----VHWSLISVDVRRRTI 236
DB 56 PEVTOFIKTSNPALIAMFLEPLDLPNKRNVFLAINDNSNQAGGTHSLVYLQDKNSF 115
QY 237 TYFDSQRTLNRRCPKHIAKYLOAEVKK-DRLDFHQGWKGYFKMNVARQNNDSCGAFVL 295
DB 116 PHYDSHSRSNSVHAKQVAEKLEAFGLGRGDKLAF-----VEEKAPAQNSYDCGMVVI 168
QY 296 QYCKHIALSQPFSFTQQ 312
DB 169 --CNTEALCQNF-FRQQ 182

RESULT 4
US-09-815-242-13640
; Sequence 13640, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815, 242
```

```

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13640
; LENGTH: 427
; TYPE: prt
; ORGANISM: Streptococcus pneumoniae
; US-09-815-242-13640

```

Query Match	5.1%	Score	90.5	DB	10	Length	427		
Best Local Similarity	20.9%	Pred. No.	0.4						
Matches	58	Conservative	51	Mismatches	114	Indels	55	Gaps	12
Qy	22	PDASILSNVCSIGDHVAQELFQGSIDLGMABEAEAPGEGKAGHSPLRHEHVTVCQSILDE	81						
Db	155	PDSCLFESTLAN-----LYPGTPLAT-----DIVGSEESISQINLTNQENFTK	198						
Qy	82	FLQTYGSLIPLSTDEVVEKLEIDFQ-----QEFSTPSRKGLVLQLTQSYORM-----P	129						
Db	199	FYKPVNMSLFLVGNFDFVERVODYFESKELKDSDFQEVAREKLFLOPKVPTDSMRMEVSSP	258						
Qy	130	GNAM-VRGFRV-----AYKRHLVTMDLGLTYGONWLNDQVMNMYGDLVMDT-----V	176						
Db	259	KLAIGVRGKREVSADCYRHILLKLLFAMFG-WTSDRFQKCYESGKIDASLSEVEV	316						
Qy	177	PEKVHFFNSFFYDK----LRTKGYDGVKRWTKNYVDIENKELLPIHL--EVHWSLSLSD	230						
Db	317	TSRPHFVWLMTDKEPVALSQFKKARNFKTLJDIITEEHLDJIIKREMGFEFFSMNSLE	376						
Qy	231	VRRRTITYDFSQRTLNRRCPKHIAKYLQAEAVKKDRDL	268						
Db	377	FIATQYDAFENGETI-----FDLPKILQ-ETLLESDVL	408						

```

RESULT 5
US-09-815-242-13443
; Sequence 13443, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essays
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

```

```

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13443
; LENGTH: 427
; TYPE: PrT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13443

```

[illegible]

```

RESULT 6
US-09-866-582-34
; Sequence 34, Application US/09866582
; Patent No. US20020127620A1
; GENERAL INFORMATION:
; APPLICANT: Witman, George B.
; APPLICANT: Fazor, Gregory J.
; APPLICANT: Rosenbaum, Joel L.
; APPLICANT: Cole, Douglas G.
; TITLE OF INVENTION: INTRAFACELLAR TRANSPORT
; FILE REFERENCE: 07917-145001
; CURRENT APPLICATION NUMBER: US/09/866,582
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/206,923
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 824
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-866-582-34

```

	Query Match	4.9%	Score 86;	DB 10;	Length 824;
	Best Local Similarity	17.6%;	Pred. No. 2.9;		
	Matches	47;	Conservative 48;	Mismatches 98;	Indels 74; Gaps 10;
Qy	67	LREEHVTCVQSILDEEFLQTGYCSLPLSTDVEVKLEIDIFQQEFSTPSRKGLVLQLIQSYQ	126		
	: : :         :   : :     : :     : :     : :				
Dd	526	LTYEKLNRLDQALPCFLLKLHAIL--RNSAEVLYQIANYIEL-MENPSQ--AIEWLMQGVVS	580		
Qy	127	RMPGNAMVRGFRVAYKRHVLTMDDLGTLYIGONWLNDQVMNMYGD--LVMTDTVPPEKVHFHN	184		

Db 581 VIPTDPQV-----LSKLGELYDREGDKSQAFYYYSYRPPCNIEVIEWLG 627  
QY 185 SFYDKLRTKGYGVKRWKNDIENKELLPIHLEHVHSLISVDVRRRTITY---FDS 241  
Db 628 AYIID-----TQFWKAQYFPRASLIQP--TQVKWQLMVASCFFRRSGNYOKALDT 676  
QY 242 QRTLNRRCPKHAQYLAQAEAVKDRDLDFHQGWKGYFKMNVARQNNDSDCGAFVLQYCKHL 301  
Db 677 YKDTHRKFFENV-----ECLRLFLVRLCTDL 701  
QY 302 ALSQPSFTQ-----QDMPKLRRIQYK 323  
Db 702 GLKDAQOYARKLRLEKMKREIQRIK 728

## RESULT 7

US-09-925-302-614  
; Sequence 614, Application US/09925302  
; Patent No. US20020044941A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT APPLICATION NUMBER: US/09/925,302  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 614  
; LENGTH: 304  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (208)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (229)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-614

Query Match 4.7%; Score 84; DB 10; Length 304;  
Best Local Similarity 18.7%; Pred. No. 1.1;  
Matches 61; Conservative 52; Mismatches 99; Indels 114; Gaps 16;

QY 65 SPLREHVTCVQSIIDEFLOTYGLSLPLSTDEWVEKLEDFQOEFSTPSRKGLVLQLIQS 124  
Db 3 NPWEKTOET-VORIL---LEPKYLLQLPGKQVRYKLSQAFNHWLKVPEDK---LQIIIE 55  
QY 125 YORMPGNAMV-----RGFRVAYKRHLVTMDDLGTLTGONWLNQVMNMGDLIV 172  
Db 56 VTEMLHNASLLDIEDNSKLRRGPPVAH-----SIYG----- 88  
QY 173 MDTVPEKVFHFFNSFY---DKLRTGY-DGVRWTKNVDIFNKELLPIH-----LEVHW 224  
Db 89 ---IPSVINSANYVFLGLEKVLTLDPDAVLFTRO-----LLELHOGQGLDIYW 136  
QY 225 S-----LISVDVRRRTITYFDSQRTL-----NRRCP 250  
Db 137 RNYTCPTPEEYKAMVLQKTGGLGLAVGLMQLFSDYKEDLPLNLTLGLFFQIIRDDVAN 196  
QY 251 KHIQYLAQAEYKVRDL-----FHQGW---KGFRKNVARQNNDSDCGAFVLQYCKH 300  
Db 197 LHSKEYSENKSCEDLTEGKFSFPIHAIWSRXSTQVQNILQRTEH---IDIKKYCVH 253  
QY 301 -LALSQPSFTQDMPKLRRIQYKEL 325  
Db 254 YLEDVGSFSEYTRNTLKELEKAYKQI 279

## RESULT 8

US-09-765-272-200  
; Sequence 200, Application US/09765272  
; Patent No. US20020061545A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/765,272  
; FILING DATE: 22-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 200:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 674 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 200:  
US-09-765-272-200

Query Match 4.7%; Score 82.5; DB 10; Length 674;  
Best Local Similarity 22.1%; Pred. No. 4.8;  
Matches 56; Conservative 35; Mismatches 79; Indels 83; Gaps 15;

QY 53 EAERPERGAGOHSPURE---EHVTCVQS---TLDEFLOT-YGSLIPLSTDEVV----- 98  
Db 44 EOENAGLSAARTGLNMGSGNVITFVDSDDWIEQDYVETLYKKIYEQADIAGVNYSFN 103  
QY 99 -----EKLED---IFQOEFSTPSRKGLVLQLIQSYQRMPGNAMVGRFV 139  
Db 104 ESEGMFYHILGDSYERKVDNVNVSIFENLYET-----QEMKSFALISAWCK 149  
QY 140 AYKRHLVFM---DDLTLGQCNWLNQVMNMGDLIVDPVKVHFHNSFFYDKLRTKYD 197  
Db 150 LYKARLEQLRFDIGKLGEDGYLNOKVY-----LLSEKVIYLNKSLYAIRKG-- 198  
QY 198 GVKR-WTKN-----VDIENKELLLI-----PI--HLEVHWSLISVDVRR-----RTIT 237  
Db 199 SLRVTWTEKMHMALVDAMSERITLLANMGYPLEKHLAVIRQMLEVSLANGQSLSDAT 258  
QY 238 YFD---SQTLLNR 247  
Db 259 YKEFEMKQRLNLQ 271

## RESULT 9

US-09-801-368-394  
; Sequence 394, Application US/09801368  
; Patent No. US20020128250A1



```
Db 653 KAFDHNLSEELGLUKNLKEDFNHVSLEDLKKASPIENFVAEKLKSDYENK----- 706
QY 115 KGLVLQIQSYQRMPCGNAMVRGRVAY-----KRHLVTMDDLGT---LYGQWLNDDQ 163
Db 707 -----MKVLSEQRS-----RIRIYVQLILDNAWEHLYTMDNLKGTGLNRGYNQKDP 756
QY 164 V-----MNVGDLVMDTVPEKVFHFNFFYD 189
Db 757 VEYKESYNLFELEFIEDIKTEAIKTSKIQFE 788

RESULT 12
US-09-815-242-5222
; Sequence 5222, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5222
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5222

Query Match 4.5%; Score 79; DB 10; Length 638;
Best Local Similarity 18.7%; Pred. No. 9.8;
Matches 56; Conservative 49; Mismatches 84; Indels 110; Gaps 13;

QY 30 NVCSIGD----HVAQLFGQSDLGMAEAEERPGERAGQ--HSPLRER----- 70
Db 59 NVSKIKDLRLGYLTQMTFNSNATVFEEMSKPFEHKRMESLIKEETDWSKHANDYDSD 118
QY 71 ----HVTVCQSILDEFQTYGSLIPLSTDEVVEKLEDFQOEFSTP-----SRKGL 117
Db 119 TVKTHMSRYESLSNQPEQLEGYQYESKIKTVLHGL-NFSEEDFNKPINDFSGGQKTRL 177
QY 118 VLQLIQSYQRMPCGNAMVRGRVAYKRHLVTMD-----DLGTYLGONWLNDOVMNMGDL 171
Db 178 AQMLLN-----EPDLLLDELPTNHLDET---TKWLEDYLRVFKGAI 216
QY 172 VMDTVPEKVFHFNFFYDKLRTKGYD----GVKRWTKNVDIFNKKELLIPIHLEVHWSLI 227
Db 217 VIIS-----HDRVFLDKIVTQYDVALGDVKRYGVNYEEF----- 251
QY 228 SVDVRRRTITY-----FDSQ-----RTLNRCPKHIAKYLOAEAVKKDRLD 268
Db 252 ----IQORDLYYOKRMOEYESQQAETKRLTFFVEKNITRASTSGMAKSRKKILEKMERID 307

RESULT 14
US-10-041-007-16
; Sequence 16, Application US/10041007
; Patent No. US20020164736A1
; GENERAL INFORMATION:
```

```
Db 252 ----IQORDLYYOKRMOEYESQQAETKRLTFFVEKNITRASTSGMAKSRKKILEKMERID 307

RESULT 13
US-09-815-242-12143
; Sequence 12143, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12143
; LENGTH: 642
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12143

Query Match 4.5%; Score 79; DB 10; Length 642;
Best Local Similarity 18.7%; Pred. No. 9.9;
Matches 56; Conservative 49; Mismatches 84; Indels 110; Gaps 13;

QY 30 NVCSIGD----HVAQLFGQSDLGMAEAEERPGERAGQ--HSPLRER----- 70
Db 59 NVSKIKDLRLGYLTQMTFNSNATVFEEMSKPFEHKRMESLIKEETDWSKHANDYDSD 118
QY 71 ----HVTVCQSILDEFQTYGSLIPLSTDEVVEKLEDFQOEFSTP-----SRKGL 117
Db 119 TVKTHMSRYESLSNQPEQLEGYQYESKIKTVLHGL-NFSEEDFNKPINDFSGGQKTRL 177
QY 118 VLQLIQSYQRMPCGNAMVRGRVAYKRHLVTMD-----DLGTYLGONWLNDOVMNMGDL 171
Db 178 AQMLLN-----EPDLLLDELPTNHLDET---TKWLEDYLRVFKGAI 216
QY 172 VMDTVPEKVFHFNFFYDKLRTKGYD----GVKRWTKNVDIFNKKELLIPIHLEVHWSLI 227
Db 217 VIIS-----HDRVFLDKIVTQYDVALGDVKRYGVNYEEF----- 251
QY 228 SVDVRRRTITY-----FDSQ-----RTLNRCPKHIAKYLOAEAVKKDRLD 268
Db 252 ----IQORDLYYOKRMOEYESQQAETKRLTFFVEKNITRASTSGMAKSRKKILEKMERID 307

RESULT 14
US-10-041-007-16
; Sequence 16, Application US/10041007
; Patent No. US20020164736A1
; GENERAL INFORMATION:
```



; APPLICANT: Matsuda, Seichi P.T.  
; APPLICANT: Schepmann, Hala G  
; TITLE OF INVENTION: Ginkgo Biloba Levopimaradiene Synthase  
; FILE REFERENCE: P02081US1  
; CURRENT APPLICATION NUMBER: US/10/041,007  
; CURRENT FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: US 60/259,881  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 16  
; LENGTH: 782  
; TYPE: PRT  
; ORGANISM: Abies grandis  
US-10-041-007-16

Query Match 4.4%; Score 78.5; DB 9; Length 782;  
Best Local Similarity 23.4%; Pred. No. 15;  
Matches 40; Conservative 32; Mismatches 54; Indels 45; Gaps 9;  
QY 43 FQSGDLGMAEEA-ERPGE-----KAGQHSPLREE-----HVTVCVQSILDEFLOTYGSIL- 89  
Db 482 FRDSGLPLFTFARERPLEFYFLVAAGTYEPQYAKCRFLFTKVACLTQVLDMDYDTYGTLD 541  
QY 90 -IPLSTDEV-----VEKLEDFIQEEFSTPSRKGVLVLQIQSYQMPGPNAMVRGFRVAY 141  
Db 542 ELKLFTEAVRRWDLSTENLPDYMKLCYQI--YYDIVHEVAWEAEKEQGRELVSPFRKGW 599  
QY 142 KHRVL-----TWDDLGLTYGQNW-----LNDQVNMNYGDLVMD 174  
Db 600 EDYLLGYVEAEWLAAYVPTLDE----YIKNGITSIGORILLSGVLMD 646

RESULT 15  
US-09-887-586A-46  
; Sequence 46, Application US/09887586A  
; Patent No. US20020094556A1  
; GENERAL INFORMATION:  
; APPLICANT: Chappell, Joseph  
; APPLICANT: No. US20020094556A11, Joseph P.  
; APPLICANT: Starks, Courtney M.  
; APPLICANT: Manna, Kathleen R.  
; TITLE OF INVENTION: SYNTHASES  
; FILE REFERENCE: 07678-025001  
; CURRENT APPLICATION NUMBER: US/09/887,586A  
; CURRENT FILING DATE: 2001-06-22  
; PRIOR APPLICATION NUMBER: 09/398,395  
; PRIOR FILING DATE: 1999-09-17  
; PRIOR APPLICATION NUMBER: 60/130,628  
; PRIOR FILING DATE: 1999-04-22  
; PRIOR APPLICATION NUMBER: 60/150,262  
; PRIOR FILING DATE: 1999-08-23  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 46  
; LENGTH: 782  
; TYPE: PRT  
; ORGANISM: Abies grandis  
US-09-887-586A-46

Query Match 4.4%; Score 78.5; DB 10; Length 782;  
Best Local Similarity 23.4%; Pred. No. 15;  
Matches 40; Conservative 32; Mismatches 54; Indels 45; Gaps 9;  
QY 43 FQSGDLGMAEEA-ERPGE-----KAGQHSPLREE-----HVTVCVQSILDEFLOTYGSIL- 89  
Db 482 FRDSGLPLFTFARERPLEFYFLVAAGTYEPQYAKCRFLFTKVACLTQVLDMDYDTYGTLD 541  
QY 90 -IPLSTDEV-----VEKLEDFIQEEFSTPSRKGVLVLQIQSYQMPGPNAMVRGFRVAY 141  
Db 542 ELKLFTEAVRRWDLSTENLPDYMKLCYQI--YYDIVHEVAWEAEKEQGRELVSPFRKGW 599  
QY 142 KHRVL-----TWDDLGLTYGQNW-----LNDQVNMNYGDLVMD 174

Db 600 EDYLLGYVEAEWLAAYVPTLDE----YIKNGITSIGORILLSGVLMD 646  
Search completed: December 16, 2002, 19:39:09  
Job time : 168 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 17:51:04 ; Search time 37 Seconds  
(without alignments)  
264.011 Million cell updates/sec

Title: US-09-848-852A-3  
Perfect score: 1772  
Sequence: 1 MYSQRFWTIARRGAHLA.....DMPKLRRIYKELCHCKLTV 332

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91.5	5.2	654	US-09-090-793-2	Sequence 2, Appl
2	91.5	5.2	661	US-08-375-709-3	Sequence 3, Appl
3	91.5	5.2	661	US-08-752-929-3	Sequence 3, Appl
4	86	4.9	727	US-09-134-001C-4067	Sequence 4067, Ap
5	83	4.7	1151	US-09-134-001C-3242	Sequence 3242, Ap
6	82.5	4.7	343	US-08-180-209B-56	Sequence 56, Appl
7	82.5	4.7	343	US-08-474-853-56	Sequence 56, Appl
8	82.5	4.7	343	US-09-166-205B-56	Sequence 56, Appl
9	82.5	4.7	343	PCT-US94-02629-56	Sequence 56, Appl
10	82.5	4.7	451	US-09-357-251-35	Sequence 35, Appl
11	82.5	4.7	674	US-08-961-083-200	Sequence 200, App
12	82	4.6	353	US-08-034-650-11	Sequence 11, Appl
13	82	4.6	353	US-08-449-015-11	Sequence 11, Appl
14	81.5	4.6	500	US-08-117-083-68	Sequence 68, Appl
15	81.5	4.6	633	US-09-041-991A-10	Sequence 10, Appl
16	81	4.6	407	US-08-989-370-6	Sequence 6, Appl
17	81	4.6	668	US-09-134-001C-4816	Sequence 4816, Ap
18	81	4.6	2987	US-08-970-269A-29	Sequence 29, Appl
19	81	4.6	2987	US-09-407-562-29	Sequence 29, Appl
20	81	4.6	3959	US-08-970-269A-30	Sequence 30, Appl
21	81	4.6	3959	US-09-407-562-30	Sequence 30, Appl
22	80.5	4.5	449	US-09-134-001C-5673	Sequence 5673, Ap
23	79.5	4.5	335	US-09-057-762-7	Sequence 7, Appl
24	79.5	4.5	335	US-08-326-119A-7	Sequence 7, Appl
25	79.5	4.5	885	US-09-342-648-9	Sequence 9, Appl
26	79.5	4.5	1010	US-09-134-001C-5178	Sequence 5178, Ap
27	78.5	4.4	436	5405943-4	Patent No. 5405943

ALIGNMENTS

RESULT 1

US-09-090-793-2  
; Sequence 2, Application US/09090793  
; Patent No. 6140486  
; GENERAL INFORMATION:  
; APPLICANT: Calgene, LLC  
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
; FILE REFERENCE: CGNE.131.01US  
; CURRENT APPLICATION NUMBER: US/09/090,793  
; CURRENT FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,650  
; EARLIER FILING DATE: 1997-06-04  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 654  
; TYPE: PRT  
; ORGANISM: Shewanella putrefaciens  
US-09-090-793-2

Query Match	5.2%	Score 91.5;	DB 4;	Length 654;	
Best Local Similarity	21.2%	Pred. No. 0.5;			
Matches	78;	Conservative	43;	Mismatches 118; Indels 129; Gaps 21;	
Qy	1	MYS---	AQRFWGTIARRGAHLAPPDASILISNVCSIGDHVAQELFGQSDLG----	MAEE 53	
Db	171	IYSHSHADHFGGA----	RGVQEMFPDVKVYGS-----	DNITKEIVDENVLAGNMSRRA 220	
Qy	54	AERPGEKAGQH-----	SPLREHVTCVQSILDEFLOTYGSLLPLSTDEVVEKLED 103		
Db	221	AYQVGATLGRHDHGVDAALGKLSGEITYVAP--	DYTLNSEGKWETLT----	IDGLEM 274	
Qy	104	IFQOEFT-----	PSRKGVLVLQIQSYQRMFG-----	NAMV 134	
Db	275	VFMDSAGEASEMITVIPSCKALWTAEI-TYQGMHNIYTLRGAKVRDALKWSKDINEMI		333	
Qy	135	RGF-----	RVAYKRHV-----	LTMDDLGTLYGQ--NWLNDQV-MNNYGDLM 173	
Db	334	NAGQDVEVLFASHAPVWGQAINDFLRLQDNYGLVHNQTLRLANDGVGIQDIDGAIQ		393	
Qy	174	DTVPEKVHFFNSFFYDKLRKTKYGVKRWTKNVD-IPENKELLIPHLHLEVHWSLISVDVR		232	
Db	394	DTIPESI-----	YKTHWTNGYHGT--YSHNAKAVYNYL-----	425	
Qy	233	RRITTFDSORTLNRCP--	KHIKAYLQ-----	AEAVKKDRLDFHQWKGCF-----KMN 280	
Db	426	----	GYFDMNPANLPLPTKQESAKFVEYMGGAADAIKRAKDDYAQGEYRFVATALNKVV		481
Qy	281	VARONDS	288		

Db 482 MAEPENDS 489

RESULT 2

US-08-375-709-3

; Sequence 3, Application US/08375709

; Patent No. 5683898

; GENERAL INFORMATION:

; APPLICANT: YAZAWA, Kazunaga

; APPLICANT: YAMADA, Akiko

; APPLICANT: KATO, Seishi

; APPLICANT: KONDO, Kiyosi

; TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid

; TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of

; TITLE OF INVENTION: Eicosapentaenoic Acid

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/375,709

; FILING DATE: 20-JAN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/178,251

; FILING DATE: 14-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 4-147945

; FILING DATE: 15-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: WEGNER, Harold C.

; REGISTRATION NUMBER: 25,258

; REFERENCE/DOCKET NUMBER: 53466/150/AAOK

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 661 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-375-709-3

Query Match 5.2%; Score 91.5; DB 1; Length 661;

Best Local Similarity 21.2%; Pred. No. 0.51;

Matches 78; Conservative 43; Mismatches 118; Indels 129; Gaps 21;

QY 1 MYS---AQRFWGTIWARGAHLAPPDASILISNVGSGIHVAQELFGSDLG----MAEE 53

Db 173 IYSHSHADHFGA---RGVQEMFDDVKVGS-----DNITKEIVDENVLGNAMSRA 222

QY 54 AERPGEKAGQH-----SPLREHVTCVQSILDBFLQTYGSLIPLSTDEVVEKLED 103

Db 223 AYQYGATLCKHDHIGVDAALGKLSKGETTYVAP---DYTLNSEGKWETLT----IDGLEM 276

QY 104 IFQEEFST-----PSRKGVLQLIQSVQRMGP-----NANV 134

Db 277 VFMDASGTEASEMITYIPSKALMTABE-LTQGMHNIYTLGAKVRDALKWSKDINEMI 335

QY 135 RGF----RVAYKRHY-----LTMDLLGTLYGQ--NWLNDQV--MNMVYGDLYM 173

Db 336 NAFQDVEVLFASHAPVWGNQAINDFLRLQRDNYGLVHNQTLRLANDGVGIQDIDAQ 395

QY 174 DTVPEKVHFFNSFFYDKLRTKGYGVKRWTKNVD--IPNKELLIPILHLEVHWSLISVDVR 232

Db 396 DTIPESI-----YKTWHTNGYHGT--YSHNAKAVYKNYL----- 427

QY 233 RRTITYFDSQRTLNRRCF--KHIAKYLQ-----APAVKKDRDLDFHQGWKGYF-----KMN 280

Db 428 ---GYFDMNPANLPLPTKQESAKFEVYMGADAAIKRAKDDYAQGEYRFVATALNKVV 483

QY 281 VARONNDS 288

Db 484 MAEPENDS 491

RESULT 3

US-08-752-929-3

; Sequence 3, Application US/08752929

; Patent No. 5798259

; GENERAL INFORMATION:

; APPLICANT: YAZAWA, Kazunaga

; APPLICANT: YAMADA, Akiko

; APPLICANT: KATO, Seishi

; APPLICANT: KONDO, Kiyosi

; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing

; TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic

; TITLE OF INVENTION: Acid

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/752,929

; FILING DATE: 20-NOV-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/375,709

; FILING DATE: 20-JAN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/178,251

; FILING DATE: 14-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 4-147945

; FILING DATE: 15-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: WEGNER, Harold C.

; REGISTRATION NUMBER: 25,258

; REFERENCE/DOCKET NUMBER: 53466/150/AAOK

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 661 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-752-929-3

Query Match 5.2%; Score 91.5; DB 1; Length 661;

Best Local Similarity 21.2%; Pred. No. 0.51;

Matches 78; Conservative 43; Mismatches 118; Indels 129; Gaps 21;





[illegible]

RESULT 9  
PCT-US94-02629-56  
; Sequence 56, Application PC/TUS9402629  
; GENERAL INFORMATION:  
; APPLICANT: King, Te-Piao  
; TITLE OF INVENTION: CLONING AND RECOMBINANT PRODUCTION OF  
; TITLE OF INVENTION: VESPID VENOM ENZYMES, SUCH AS PHOSPHOLIPASE AND  
; TITLE OF INVENTION: HYALURONIDASE, AND IMMUNOLOGICAL THERAPIES BASED THEREON  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/02629  
; FILING DATE: 10-MAR-1994  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/180,209  
; FILING DATE: 11-JAN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/031,400  
; FILING DATE: 11-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-074 PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 56:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 343 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEITICAL: NO  
; ANTI-SENSE: NO  
PCT-US94-02629-56

```
Db      64 KDPNGVVARNGVPQGNLTKHLQVFDRDLHNLINQIDPKSFGPGVGVIDFESWR----- 115-  
           |::| : : : | : : : | : : ||  
  
Qy     212 ELLIPITHLEVHWSS-----LISDVRRRTITTFDSQRTLNRRCRPHKIAKYLAQAEEAVKKD 265  
           || : ::||:||||: :|| : :||:  
  
Db     116 -----PIFRQ-NWASLPYKKLSVEVVVRREHPWDQR-VQEAKRKREEKGQ----- 161  
  
Qy     266 RLDFHQGWKFKNVARQNNDSCGAFVLQYCCKHALSQPF-----SFTQQDDMPKL 317  
           |:| : | : :| : : :| :| : ||:  
  
Db     162 -LFMEETLKAARKMRPA-----ANWGYYAYPYCYNLTNPQSACQBATTMOENDKM 211  
  
RESULT 10  
US-09-357-251-35  
; Sequence 35, Application US/09357251  
; Patent No.. 6271441  
; GENERAL INFORMATION:  
; APPLICANT: Falco, S. Carl  
; APPLICANT: Famodu, Layo O.  
; APPLICANT: Orozco, Buddy  
; APPLICANT: Schwaber, James S.  
; TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase  
; FILE REFERENCE: BB-1193  
; CURRENT APPLICATION NUMBER: US/09/357,251  
; CURRENT FILING DATE: 1999-07-20  
; EARLIER APPLICATION NUMBER: 60/093,530  
; EARLIER FILING DATE: July 21, 1998  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 35  
; LENGTH: 451  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-357-251-35
```

```

RESULT 11
; US-08-961-083-200
; Sequence 200, Application us/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
;

```

COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 200:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 674 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-083-200

Query Match 4.7%; Score 82.5; DB 4; Length 674;  
Best Local Similarity 22.1%; Pred. No. 4.9;  
Matches 56; Conservative 35; Mismatches 79; Indels 83; Gaps 15;  
QY 53 EAEPEGEKAGQSPURE---EHVTCVQS---ILDEFLOT-YGSLIPLSTDEVV----- 98  
Db 44 EQENAGLSAARTGLNMGSGYVTFVDSDDTIEQDYVETLYKRIYEQADIAGVNYSFN 103  
QY 99 -----EKLED---IFQOEFTSPSRKGLVLQLIQSYVQRMGNAMVGRGV 139  
Db 104 ESEMEYFHLGDSYKERYDNVSIFENDYET-----QEMKSFALISAMGK 149  
QY 140 AYKRHRVLTN--DGLTLYGQNLNDQVMNMGDLVMDTVPEKVFHFNFPYDKLRTKGYD 197  
Db 150 LYKARLEQRLDICKLGEDGYLNQKV-----LLSERVIYNKSLIAYIRKG-- 198  
QY 198 GVKR-WTKN-----VDIFNKELLI-----PI--HLEVHWSLISVDVRR-----RTIT 237  
Db 199 SLRVWTERKWMHALVDAMSERITLLANMGYPLEKHLAVYRQMLEVSLANGQASGLSDTAT 258  
QY 238 YFD---SORTLNR 247  
Db 259 YKEFEMKQRLLAQ 271

RESULT 12  
US-08-034-650-11  
; Sequence 11, Application US/08034650  
; Patent No. 5641671  
; GENERAL INFORMATION:  
; APPLICANT: BOS, Jannetje W.  
; APPLICANT: FRENKEN, Leon G.  
; APPLICANT: VERRIPS, Cornelis T.  
; APPLICANT: VISSER, Christiaan  
; TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L. Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA

ZIP: 20036-5601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/034,650  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/727,235  
FILING DATE: 03-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kokulis, Paul N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: PNK/5970/91731  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 353 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-034-650-11  
Query Match 4.6%; Score 82; DB 1; Length 353;  
Best Local Similarity 20.8%; Pred. No. 2;  
Matches 48; Conservative 37; Mismatches 94; Indels 52; Gaps 9;  
QY 46 SDLGMAEAEAPGEGKAGHSPL-----REEHVTVCVQSILDEFLOTYGSILPLSTDEVV 98  
Db 69 AEAPMLPALPALGALAGSHAPRLPLAAGRLARTRAVREFFDYCLTAQOGLTPAALDALV 128  
QY 99 EKLEDIFQOEFTSPSRKGL-VLQLIQSY-----QRMGNAMVGRFVAYKRHRVLTMDDLGT 153  
Db 129 RR-ETAQLDQSPAQAEALGVWRRYRAYFDALAQPLPGDGLDPAAMQLALDQRAA 187  
QY 154 LYQWNLNDQVMNMGDLVMDTVPEKVFHFNFPYDKLRTKGYDGVKRWTKNVDIFNKEL 213  
Db 188 -----LADRTLGEAE-----PFFGDEQRQRH-----LERIRLAND-- 220  
QY 214 LLPIHLEVHWSLISVDVRRRTITYFDSQRLNRRCPKHIKYLQAEAVKK 264  
Db 221 -----TTLSPEQKAARLAALDAQLTDPDERA-QOAAALHAQODAVTK 259

RESULT 13  
US-08-449-015-11  
; Sequence 11, Application US/08449015  
; Patent No. 5804409  
; GENERAL INFORMATION:  
; APPLICANT: BOS, Jannetje W.  
; APPLICANT: FRENKEN, Leon G.  
; APPLICANT: VERRIPS, Cornelis T.  
; APPLICANT: VISSER, Christiaan  
; TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L. Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS



```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,015
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/727,235
; FILING DATE: 03-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kokulis, Paul N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: PNK/5970/91731
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-449-015-11
```

```
Query Match 4.6%; Score 82; DB 1; Length 353;
Best Local Similarity 20.8%; Pred. No. 2;
Matches 48; Conservative 37; Mismatches 94; Indels 52; Gaps 9;

QY 46 SLDGMAEAEERGEKAGQHSPL-----REEHVTCSILDEFLQTYGSLIPLSTDEVV 98
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 69 AEAAMPPLPAALPGALAGSHAPRLPAAGGLARTRAVREFFDYCLTAQELTPAALDALV 128
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 99 EKLEDFIQEETSPSKGI-VLQLQSY----QRMFGNMGVFRVAYKRHRVLTWDLLGT 153
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 129 RR-ETAAQLDGSPAQAALGWRRTFRAYFDALAQPGDGLGDKLDPAAQLADQRAA 187
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 154 LYQGNLNDQNMNMYGDLVMDTVPEKVHFFNSFFYDKLRTKGYDGVKRWTKNVDIFNKEL 213
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 188 -----LADRTLGEAAE-----PFFGDEQRQRHD-----LERIRIAND-- 220
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 214 LLIPHTLHVHWSLISVDVRRRTITFYDSORTLNRRCPKHIAKYLOAEAVKK 264
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 221 -----TTLSPEQKAARLAALDAQLTPDERA-QQAALHAQQDAVTK 259
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 14
US-08-117-083-68
; Sequence 68, Application US/08117083
; Patent No. 5719054
; GENERAL INFORMATION:
; APPLICANT: Boursnell, Michael E.
; APPLICANT: Inglis, Stephen C.
; APPLICANT: Munro, Alan J.
; TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human
; TITLE OF INVENTION: Papilloma Virus Proteins
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Walter H. Dreger
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,083
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-58783
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..500
; OTHER INFORMATION: /note= "Xaa refers to stop codon in
; OTHER INFORMATION: the open reading frame."
; US-08-117-083-68

Query Match 4.6%; Score 81.5; DB 1; Length 500;
Best Local Similarity 19.8%; Pred. No. 4;
Matches 51; Conservative 37; Mismatches 86; Indels 83; Gaps 9;

QY 82 FLQTYGSLIPLSTDEVVEKLEDFIQ-----QEFSTP-----SRKGLVLQLIQS 124
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 YNKTGSCIQDMSBALEYLSLKSGGWSRPPEQFEYDPGVEDTESIERLVEEFFNR 123
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 125 YQMPGNAMVRCFRVAYKRHRVLTMDLLGTLYGQNNLNDQNMNMYGDLVMDTVPEKVHFFN 184
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 124 SELQAGESVKFGNSINVKHTSVSAKQLRTIRQQ-----LPLYSHLLP----- 166
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 185 SFYDKLRTKGYDGVKRWTKNVDIFNKELLIIPI-HLEVHWSL----- 226
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 167 -----TORVDICSLIIIIHTKXLDLSRWYMTTSDWXRMIKXSISKDF 208
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 227 ---ISVDVRRRTITFYDSORTLN--RRCPKH---IAKYLOAEAVKKDRLDHOGW----- 273
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 209 VLCISVRKRRTSSRVDSRYINLGMRLPRHTCALSKWKDAVVLCIQGNHGILMVS 268
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 274 KGYFKMNVARQNNDSDC 290
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 269 RSLQMGNGCHIXNFNC 285
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
US-09-041-991A-10
; Sequence 10, Application US/09041991A
; Patent No. 6107278
; GENERAL INFORMATION:
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Muller-Cohn, Judy
; TITLE OF INVENTION: Toxins Active Against Pests
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,991A
; FILING DATE: 13-MAR-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
```

; NAME: Sanders, Jay M.  
; REGISTRATION NUMBER: 39,355  
; REFERENCE/DOCKET NUMBER: MA-709  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (352) 375-8100  
; TELEFAX: (352) 372-5800  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 633 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-09-041-991A-10

Query Match 4.6%; Score 81.5; DB 3; Length 633;  
Best Local Similarity 23.2%; Pred. No. 5.7;  
Matches 52; Conservative 33; Mismatches 96; Indels 43; Gaps 11;  
QY 24 ASILISNVCS-IGDHVAQEL----FGSDLGMAEEAEPGEKAGOHSPLEEHVTCVQSI 78  
DB 58 ASFLKKGSLGKRLSELRLNLIFFSGSTNLMEDILRETEKF-LNQKLTDTLSRVNAE 116  
QY 79 L-----DEFLQYGSGLIPLSTDEVVEKLEDFQOEFSTPSRKGLVLQLIQS 124  
DB 117 LTGLQANVEEFNRQVDNFLNPNRNAVPLSTSSVNTMQQLNRLSQFQMGVQQLL--- 173  
QY 125 YQRMPCGNMVRGFRVAYKRHV-LTMDDLG-----TLYG-QNWLNQVMNMYGDLVMDTVPE 178  
DB 174 ---LPLFAQAANTHLSYIRDVILNAEEWGISATLFTYQNHLENTYTRD-YSNYCIDTYQT 229  
QY 179 KVHFENSEFYDKLRTKGY-----DGVKRWTKNVDIFNKELLLI 216  
DB 230 AFRGLNRIHMLFEPTYMFLNVEFYVSIWS-----LFKYQSLLV 269

Search completed; December 16, 2002, 19:28:42  
Job time : 39 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 19:28:00 ; Search time 899 Seconds  
(without alignments)  
49.209 Million cell updates/sec

Title: US-09-848-852A-3

Perfect score: 1772

Sequence: 1 MYSQRFWGTIWARGAHLA.....DMPKLRQYKELCHCKLV 332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1667	94.1	371	21	AAV50911 Human fetal brain
2	1667	94.1	371	22	AAW39269 Human polypeptide
3	1667	94.1	438	22	AAW41055 Human polypeptide
4	1667	94.1	438	22	AAW41056 Human polypeptide
5	1667	94.1	568	22	AAW31977 Amino acid sequenc
6	1558.5	88.0	354	22	AAW39270 Human polypeptide
7	1470	83.0	309	23	ABP41139 Human ovarian anti
8	781.5	44.1	270	22	ABW25617 Human protein sequ
9	614.5	34.7	119	22	AAO04440 Human polypeptide
10	490	27.7	191	22	AAW92493 Human protein sequ

11	456	25.7	643	22	AAW31973 Amino acid sequenc
12	393	22.2	509	22	AAW31976 Amino acid sequenc
13	393	22.2	589	22	AAW94011 Human stomach canc
14	393	22.2	589	22	AAW95254 Human protein sequ
15	311	17.6	468	22	ABW59432 Drosophila melanog
16	307	17.3	1513	22	ABW63592 Drosophila melanog
17	280	15.8	51	21	AAW34355 Human secreted pro
18	260.5	14.7	446	21	AAW45771 Arabidopsis thalia
19	260.5	14.7	489	21	AAW45770 Arabidopsis thalia
20	258.5	14.6	305	21	AAW45772 Arabidopsis thalia
21	257.5	14.5	446	21	AAW25055 Arabidopsis thalia
22	257.5	14.5	489	21	AAW25054 Arabidopsis thalia
23	255.5	14.4	305	21	AAW25056 Arabidopsis thalia
24	249.5	14.1	198	20	AAW86186 Human EST product
25	249.5	14.1	489	20	AAW86184 Arabidopsis ESDA (
26	205	11.6	209	20	AAW86185 Rice EST product s
27	201.5	11.4	566	23	ABW90943 Herbicidally activ
28	190	10.7	548	21	AAW77935 A. thaliana enviro
29	169.5	9.6	1148	22	ABW64875 Drosophila melanog
30	158.5	8.9	98	22	AAE01755 Human gene 11 enco
31	154	8.7	560	22	ABW71015 Drosophila melanog
32	143	8.1	362	21	AAW91646 Human secreted pro
33	143	8.1	1112	22	AAW78297 Human SUMO-1 SEQ 1
34	143	8.1	1112	22	AAW40358 Human polypeptide
35	143	8.1	1115	22	ABG14039 Novel human diagno
36	125	7.1	368	22	ABW94888 Human protein sequ
37	122.5	6.9	238	23	ABW97801 Human secretory po
38	119.5	6.7	223	23	ABW04392 Human protein phos
39	116	6.5	162	22	ABW07019 Novel human diagno
40	112	6.3	51	21	AAW34354 Gene 4 human secre
41	110	6.2	215	22	ABW63890 Drosophila melanog
42	107.5	6.1	272	23	ABW53442 Lactococcus lactis
43	106	6.0	2464	22	AAW78919 Human protein SEQ
44	104.5	5.9	272	23	ABW54137 Lactococcus lactis
45	104.5	5.9	272	23	ABW55429 Lactococcus lactis

ALIGNMENTS

RESULT 1

AAV50911  
ID AAV50911 standard; Protein; 371 AA.

XX AAV50911;

AC AAV50911;

XX 10-MAR-2000 (first entry)

XX Human fetal brain cDNA clone vb7\_1 derived protein #1.

XX Human; secreted protein; treatment: nutritional activity; cytokine;  
KW cell proliferation; cell differentiation; hematopoiesis regulation;  
KW tissue growth; activin; inhibin; chemotactic; chemokinetic; hemostatic;  
KW thrombolytic; anti-inflammatory; invasion suppressor; tumor inhibition;  
KW gene therapy.  
XX Homo sapiens.  
OS Homo sapiens.  
XX WO9955721-A1.  
PN WO9955721-A1.  
XX 04-NOV-1999.  
PD 04-NOV-1999.  
XX 23-APR-1999; 99WO-US08504.  
PF 23-APR-1999; 99WO-US08504.  
XX 24-APR-1998; 98US-0082904.  
PR 11-JUN-1998; 98US-0088994.  
PR 12-JUN-1998; 98US-0089278.  
PR 02-JUL-1998; 98US-0091647.  
PR 24-AUG-1998; 98US-0097639.  
PR 22-APR-1999; 99US-0097639.  
XX (ALPH-) ALPHAGENE INC.

PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;  
 XX WPI; 2000-052801/04.  
 DR N-PSDB; AA243782.  
 XX  
 XX New polynucleotides encoding secreted human proteins, derived from  
 PT human fetal brain, adult skin, adult brain, adult heart, adult thymus  
 PT and adult aorta cDNA libraries.  
 XX  
 XX Claim 21a; Page 223-225; 282pp; English.  
 XX  
 CC This invention describes novel human secreted proteins which are encoded  
 CC by polynucleotides obtained from fetal brain, adult skin, adult brain,  
 CC adult heart, adult thymus and adult aorta cDNA libraries. The  
 CC polynucleotides and proteins are predicted to have biological activities  
 CC which would make them suitable for treating, preventing or ameliorating  
 CC medical conditions in humans and animals, although no supporting data  
 CC is given. Suggested activities include nutritional activity, cytokine  
 CC and cell proliferation/differentiation activity, immune stimulating  
 CC (e.g. as vaccines) or suppressing activity, hematopoiesis regulating  
 CC activity, tissue growth activity, activin/inhibin activity,  
 CC chemotactic/chemokinetic activity, hemostatic and thrombolytic activity,  
 CC receptor/ligand activity, anti-inflammatory activity, cacharin/tumor  
 CC invasion suppressor activity, and tumor inhibition activity. The  
 CC polynucleotides are also stated to be useful for gene therapy.  
 CC AA50905-Y50947 represent the secreted proteins described in the method  
 CC of the invention which are encoded by the polynucleotides represented in  
 CC AA243777-243808.  
 XX  
 XX Sequence 371 AA;

Query Match 94.1%; Score 1667; DB 21; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 7e-157;  
 Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGQHSPLREHVTVCVOSI 78  
 DB 58 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGQHSPLREHVTVCVOSI 117  
 QY 79 LDEFLOTYSGLIPLSTDVEVEKLEDFQOEFSTPSRKGVLVLQILOSQYORMPGNAMVRGFR 138  
 DB 118 LDEFLOTYSGLIPLSTDVEVEKLEDFQOEFSTPSRKGVLVLQILOSQYORMPGNAMVRGFR 177  
 QY 139 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 198  
 DB 178 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 237  
 QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 258  
 DB 238 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 297  
 QY 259 AEAVKKDLRDLFHQGWKGYFKMVARQNNDSDCGAFVLOYCKHLALSOPFTQDDMPKLR 318  
 DB 298 AEAVKKDLRDLFHQGWKGYFKMVARQNNDSDCGAFVLOYCKHLALSOPFTQDDMPKLR 357  
 QY 319 ROIYKELCHCKLTV 332  
 DB 358 ROIYKELCHCKLTV 371

## RESULT 2

AAAM39269  
 ID AAM39269 standard; Protein; 371 AA.

XX  
 AC AAM39269;

XX  
 DT 22-OCT-2001 (first entry)

XX  
 DE Human polypeptide SEQ ID NO 2414.

XX  
 KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW anyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200153312-A1.  
 PN  
 XX 26-JUL-2001.  
 PD  
 XX 26-DEC-2000; 2000WO-US34263.  
 PF  
 XX 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR N-PSDB; AAI58425.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 XX Example 4; SEQ ID NO 2414; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX Sequence 371 AA;  
 Query Match 94.1%; Score 1667; DB 22; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 7e-157;  
 Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGQHSPLREHVTVCVOSI 78  
 DB 58 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGQHSPLREHVTVCVOSI 117  
 QY 79 LDEFLOTYSGLIPLSTDVEVEKLEDFQOEFSTPSRKGVLVLQILOSQYORMPGNAMVRGFR 138  
 DB 118 LDEFLOTYSGLIPLSTDVEVEKLEDFQOEFSTPSRKGVLVLQILOSQYORMPGNAMVRGFR 177  
 QY 139 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 198  
 DB 178 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 237  
 QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 258  
 DB 238 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 297







CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
CC vaginitis), immune disorders (e.g., congenital and acquired  
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
CC respiratory disorders, neurological disorders, gastrointestinal disorders  
CC and urinary system disorders. Ovarian antigen polypeptides and  
CC polynucleotides may also be used in screening for compounds which  
CC modulate ovarian antigen expression or activity. The polynucleotides may  
CC further be used for gene therapy, chromosome mapping, in the  
CC identification of individuals and in forensic analysis, and the  
CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents a human ovarian antigen of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 309 AA;

Query Match 83.0%; Score 1470; DB 23; Length 309;  
Best Local Similarity 97.2%; Pred. No. 2e-137;  
Matches 276; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 49 GMAEAEERPKAGOSPLREHVTVCQSILDEFLOTYSGLIPLSTDEVVEKLEDFQOE 108  
DB 26 GHRREGRAAGEKAGOSPLREHVTVCQSILDEFLOTYSGLIPLSTDEVVEKLEDFQOE 85  
QY 109 FSTPSKGLVLQIQSYORMPGNMVRGFRVAYKRHLVTMDLGTLYGQNLNDQVMNNY 168  
DB 86 FSTPSKGLVLQIQSYORMPGNMVRGFRVAYKRHLVTMDLGTLYGQNLNDQVMNNY 145  
QY 169 GDLYVMDTPEKVFHNSFFYDKLRGKGVGKRWTKNVDIFNKELLIPILHLEVHWSLIS 228  
DB 146 GDLYVMDTPEKVFHNSFFYDKLRGKGVGKRWTKNVDIFNKELLIPILHLEVHWSLIS 205  
QY 229 VDVRRTITYFDSQTLNRCPKHAIAKYLQAEAVKDRDLDFHQGWKGYFKMVARQNDS 288  
DB 206 VDVRRTITYFDSQTLNRCPKHAIAKYLQAEAVKDRDLDFHQGWKGYFKMVARQNDS 265  
QY 289 DCGAFVLYQCKHLALSQPSFSTQDMPKLRROIYKELCHCKLTV 332  
DB 266 DCGAFVLYQCKHLALSQPSFSTQDMPKLRROIYKELCHCKLTV 309

RESULT 8  
AAM25617  
ID AAM25617 standard; Protein: 270 AA.

XX AC AAM25617;

XX DT 16-OCT-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:1132.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiac; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;  
KW dermatological; anti-allergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmune;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; diabetes; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder.

XX OS Homo sapiens.

XX WO200153455-A2.  
PN 26-JUL-2001.  
PD 22-DEC-2000; 2000WO-US35017.  
PF 23-DEC-1999; 99US-0471275.  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PA (HYSE-) HYSEQ INC.  
Tang YT, Liu C, Drmanac RT;  
WPI: 2001-457603/49.  
DR N-PSDB; AAH999558.  
XX Isolated human polynucleotides encoding polypeptides, useful for the  
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection .  
XX Claim 20; Page 235; 1217pp; English.

XX AAM259166 to AAH99904 encode the human proteins given in AAM25225 to  
CC AAM25963. The proteins can have activities based on the tissues and  
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;  
CC antiulcer; osteopathic; dermatological; anti-allergic; antiasthmatic;  
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
CC encoding them can be used in gene therapy, antisense therapy and vaccine  
CC production. The proteins and polynucleotides are useful for screening for  
CC agonists or antagonists of a protein and for the treatment and diagnosis  
CC of disorders associated with the activity of a protein e.g. inflammation,  
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.

XX Sequence 270 AA;

Query Match 44.1%; Score 781.5; DB 22; Length 270;  
Best Local Similarity 54.0%; Pred. No. 4.2e-69;  
Matches 148; Conservative 54; Mismatches 65; Indels 7; Gaps 4;

QY 58 GEKAGOSPLREHVT-CVQSILDEFLOTYSGLIPLSTDEVVEKLEDFQOEFTPSRKG 116  
DB 1 GKNQKASPDVDEQLSVCLSGFLDEVMKKYGLVPLSEKEVLGRKLVDFNEDFS--NRKP 58  
QY 117 LVLLQIQSYORMPGNMVRGFRVAYKRHLVTMDLGTLYGQNLNDQVMNNYGDLYVMDTV 176  
DB 59 FINEITNWRARHOKC---NFRIFYNKIMLMDLTLATLDGQNLNDQVINMYGELIMDAV 115  
QY 177 PEKVHFFNSFFYDKLRGKGVGKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTI 236  
DB 116 PDKVHFFNSFFHRLQVLTGKYGNGVRWTKKVDLFKKSLLLIPILHLEVHWSLITVLSNRII 175  
QY 237 TYFDSQTLNRCPKHAIAKYLQAEAVKDRDLDFHQGWKGYFKMVARQNDSDCGAFVLQ 296  
DB 176 SFYDSQGHFFKFCVENIRKYLLEAREKNLNL--QGMOTAVTKCIPOQKNDSDCGVFLQ 234  
QY 297 YCKHLALSQPSFSTQDMPKLRROIYKELCHCKL 330  
DB 235 YCKLALQKPFQSQEDMPVRVKRIYKELCBRL 268

RESULT 9







CC for screening for compounds to treat the disease. They can also be used  
CC for predicting micro-metastases. The gene can predict peritoneal  
CC dissemination.

QY 13 ARGAHLAPPDASILISNVCISGIDHVAQELFOGSDLGMAEAEAPGKAGOHSPLEEHV 72  
Db 231 SRRGYQLEP-----DLSEVSARLRLGS-----GSLRLRRK-- 262  
QY 73 TCQVSIIDLEFQYGLSILPSTDEVVEKLEDFIQOEFSTPSRKLVLQLIOSYQMPGNA 132  
Db 263 --VSIITKKNCSGKERDRRTDDELTEDM-EKEISNALGHGQPODEILSS----- 311  
QY 133 MVRGFRVAYKRRLVMTDDGLTLYQGNWLNDOVMNMYGDLVMDTVPEK----VHFFNSFFY 188  
Db 312 -----AFKLRI--TRGDIHTLKNYHWNDEVINFYNNLLVERNKKOGYPALHVSTFFY 363  
QY 189 DKLRTKGYDGKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITTFDSORTLNR 248  
Db 364 PKLKSOGYQAVKRWTKGVNLFQEIIILVPIHRKVHWSLVLDLKKCKLYLDSMGQKGR 423  
QY 249 CPHIAKYLAQAEVKKDRLDF-----HOGWKGYFKMNVARQNNDSCGAFVLYQCKHLA 302  
Db 424 ICEILLYQLODESKTKRNSDLNLEWTHSMKPH---EIPQOLNGSDCGMFTCKYADYIS 480  
QY 303 LSOPFSFTQODMPKLRRIQYKELCHCKL 330  
Db 481 RDXPITFTQHOMPLFRKKMWEILHQOL 508

RESULT 13

AAM94011  
ID AAM94011 standard; Protein; 589 AA.

XX AC AAM94011;  
XX AC Human stomach cancer expressed polypeptide SEQ ID NO 92.  
XX DT 13-NOV-2001 (first entry)  
XX DE Human stomach cancer expressed polypeptide SEQ ID NO 92.  
XX KW Human; stomach cancer; marker; screening; micro-metastasis;  
XX KW peritoneal dissemination.

XX OS Homo sapiens.

XX PN W020109317-A1.

XX PD 08-FEB-2001.

XX PF 28-JUL-2000; 2000WO-JP05063.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 18-OCT-1999; 99US-0159590.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 17-FEB-2000; 2000US-0183322.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Aburatani H;

XX PI Kodama T, Midorikawa Y;

XX PS WPI; 2001-570287/64.

XX DR N-PSDB; AAI93888.

XX XX New Stomach cancer-associated genes, useful as markers in blood tests

XX PT for screening for the early stages of the disease -

XX PS Claim 1; Page 167-169; 242pp; Japanese.

XX XX The invention relates to stomach cancer-expressed genes

CC (AAI93842-AAI93917) and the encoded proteins (AAM93967-AAM94039). The

CC genes can be used as markers in blood tests for screening for the early

CC stages of the disease. The proteins and peptides can be used as targets

XX

CC for screening for compounds to treat the disease. They can also be used  
CC for predicting micro-metastases. The gene can predict peritoneal  
CC dissemination.

QY 13 ARGAHLAPPDASILISNVCISGIDHVAQELFOGSDLGMAEAEAPGKAGOHSPLEEHV 72  
Db 311 SRRGYQLEP-----DLSEVSARLRLGS-----GSLRLRRK-- 342

QY 73 TCQVSIIDLEFQYGLSILPSTDEVVEKLEDFIQOEFSTPSRKLVLQLIOSYQMPGNA 132  
Db 343 --VSIITKKNCSGKERDRRTDDELTEDM-EKEISNALGHGQPODEILSS----- 391

QY 133 MVRGFRVAYKRRLVMTDDGLTLYQGNWLNDOVMNMYGDLVMDTVPEK----VHFFNSFFY 188  
Db 392 -----AFKLRI--TRGDIHTLKNYHWNDEVINFYNNLLVERNKKOGYPALHVSTFFY 443

QY 189 DKLRTKGYDGKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITTFDSORTLNR 248  
Db 444 PKLKSOGYQAVKRWTKGVNLFQEIIILVPIHRKVHWSLVLDLKKCKLYLDSMGQKGR 503

QY 249 CPHIAKYLAQAEVKKDRLDF-----HOGWKGYFKMNVARQNNDSCGAFVLYQCKHLA 302  
Db 504 ICEILLYQLODESKTKRNSDLNLEWTHSMKPH---EIPQOLNGSDCGMFTCKYADYIS 560

QY 303 LSOPFSFTQODMPKLRRIQYKELCHCKL 330  
Db 561 RDXPITFTQHOMPLFRKKMWEILHQOL 588

RESULT 14  
AAB95254  
ID AAB95254 standard; Protein; 589 AA.

XX AC AAB95254;

XX XX 26-JUN-2001 (first entry)

XX XX Human protein sequence SEQ ID NO:17423.

XX XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX PT full-length cDNAs defined in the specification, and for the detection

XX PT and/or diagnosis of the abnormality of the proteins encoded by the

XX PT full-length cDNAs -

XX

PS Claim 8; SEQ ID 17423; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-d primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX

SQ Sequence 589 AA;

Query Match 22.2%; Score 393; DB 22; Length 589;

Best Local Similarity 29.6%; Pred. No. 5.2e-30;

Matches 97; Conservative 64; Mismatches 107; Indels 60; Gaps 9;

QY 13 ARGAHAPPDASILISVCSIGDHVACELFGQSLGMAEEAERGERAGOHSPLEEHV 72

DB 311 SRGQLEP-----DLSEVSARLRGS-----GNGLLRRK-- 342

QY 73 TCVCISILDEFLQYGLSLPLSTDEYVEKLEDFIQOEFSTPSRKGVLQIQSYQRMPCNA 132

DB 343 --VSIETKEKNCSEKDRRTDDELLETDW-EKEISNALGHGQDEILSS----- 391

QY 133 MYRGFRVAYKRVHLMDDGLTYGONWLDQVMNMYGDLVMDTVPEK-----VHFNSEFF 188

DB 392 -----AFKLRI-TRGDHITLKNVHNLNDEVINFYMLLVERNKKQGYPALHVFSTFF 443

QY 189 DKLRKGVGKRWTKNVDIENKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRR 248

DB 444 PKLSSGGQAVKRWTKGVNLEFQEIILPIHRKVHWSLVLDLRKKLKYLDMSGCKGHR 503

QY 249 CPKHIAKYLAQAVKDKRLDF-----HOGWKGYFKMNVARQNNDSCGAFVLYQCKHLA 302

DB 504 ICEILLQLQDESKTKRNSDLNLEWTHHSMKPH---EIPQQLNGSDCGMFTCKYADVIS 560

QY 303 LSQPFSTQQDMPKLRRIYKELCHCKL 330

DB 561 RDKPITFTQHOMPLFRKKMNVWEILHQOL 588

RESULT 15

ABB59432

ID ABB59432 standard; Protein; 468 AA.

XX

AC ABB59432;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 5088.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

XX

OS Drosophila melanogaster.

XX

FN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL03535.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

PS Disclosure; SEQ ID NO 5088; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 468 AA;

Query Match 17.6%; Score 311; DB 22; Length 468;

Best Local Similarity 28.5%; Pred. No. 5.2e-22;

Matches 70; Conservative 55; Mismatches 85; Indels 36; Gaps 6;

QY 89 LIPLSTDEVVEKLEDFIQOEFSTPSRKGVLQIQSYQRMPCGNAMVRCFRVAYKRHV-LT 147

DB 245 LIPL-TEEHDRFRNEITQDDKST-----EIIKFNLHIT 277

QY 148 MDDLGLTYGONWLDQVMNMYGDLVMDTVPEK-----VHFNSEFFYDKLRTKYGDKVKR 201

DB 278 TEDICTFNGKWLNDVINFYMSLLTERSEKRSGLVPATYAINTFVPRLLQAGHAGIKR 337

QY 202 WTKNVDIFNKELLIPILHLE-VHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLAQAE 260

DB 338 WTRKVDLFSKDIIPVPVHCNGVHWCMAIIHLRNKTIIRYDSKGNRPVLDALKYLEE 397

QY 261 AVKDKRLDFH-QGKWGYFKMNVARQNNDSCGAFVLYQCKHLALSQPFSTQQDMPKLR 319

DB 398 SIFPKKQFDTSDFVIESVQNIPRLQDSCGIFSCMFAEYITCDVPITFTQSEMFLFRK 457

QY 320 QIYKEL 325

DB 458 KMALEI 463

Search completed: December 16, 2002, 19:54:29

Job time : 900 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2002, 20:11:20 ; Search time 98 Seconds

(without alignments)  
1339.725 Million cell updates/sec

Title: US-09-848-852A-3

Perfect score: 1772

Sequence: 1 MYSQRFWTGIIWARRGAHLA.....DMPKLRRIYKELCHCKLTIV 332

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 355320 seqs, 197730502 residues

Total number of hits satisfying chosen parameters: 710640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2\_1/USPTO\_spool/US09848852/runat\_16122002\_132039\_19075/app\_query.fasta\_1.519  
-DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09848852\_@CGN\_1.1.49\_@runat\_16122002\_132039\_19075  
-NCPU=6 -ICPU=3 -NO\_XLPHY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications\_NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	1772	100.0	1991	10	US-09-848-852A-4	Sequence 4, Appli
c 2	704	39.7	432	10	US-09-983-965-835	Sequence 835, App
c 3	661.5	37.3	503	10	US-09-777-564-1710	Sequence 1710, Ap
4	222.5	12.6	1866	10	US-09-856-247A-1	Sequence 1, Appli

5	162	9.1	435	10	US-09-867-701-4735	Sequence 4735, Ap
6	144	8.1	5540	10 <td>US-09-960-253-74</td> <td>Sequence 74, Appl</td>	US-09-960-253-74	Sequence 74, Appl
7	130.5	7.4	264	10 <td>US-09-878-574-7902</td> <td>Sequence 7902, Ap</td>	US-09-878-574-7902	Sequence 7902, Ap
8	122.5	6.9	639	12 <td>US-10-008-461-1</td> <td>Sequence 1, Appli</td>	US-10-008-461-1	Sequence 1, Appli
9	115.5	6.5	100	10 <td>US-09-728-445-223</td> <td>Sequence 223, App</td>	US-09-728-445-223	Sequence 223, App
10	110	6.2	263	10 <td>US-09-294-093B-3098</td> <td>Sequence 3098, Ap</td>	US-09-294-093B-3098	Sequence 3098, Ap
11	108	6.1	4835	10 <td>US-09-956-004-58</td> <td>Sequence 58, Appl</td>	US-09-956-004-58	Sequence 58, Appl
c 12	103	5.8	547	10 <td>US-09-924-401-19</td> <td>Sequence 19, Appl</td>	US-09-924-401-19	Sequence 19, Appl
13	96	5.4	472	10 <td>US-09-783-590-7988</td> <td>Sequence 7988, Ap</td>	US-09-783-590-7988	Sequence 7988, Ap
14	93.5	5.3	2529	10 <td>US-09-925-302-171</td> <td>Sequence 171, App</td>	US-09-925-302-171	Sequence 171, App
15	90.5	5.1	1284	10 <td>US-09-815-242-9542</td> <td>Sequence 9542, Ap</td>	US-09-815-242-9542	Sequence 9542, Ap
16	90	5.1	2395	9 <td>US-10-063-547-139</td> <td>Sequence 139, App</td>	US-10-063-547-139	Sequence 139, App
17	90	5.1	2395	12 <td>US-10-036-342-44</td> <td>Sequence 44, Appl</td>	US-10-036-342-44	Sequence 44, Appl
18	90	5.1	2395	12 <td>US-10-006-867-139</td> <td>Sequence 139, App</td>	US-10-006-867-139	Sequence 139, App
19	89.5	5.1	1284	10 <td>US-09-815-242-9345</td> <td>Sequence 9345, Ap</td>	US-09-815-242-9345	Sequence 9345, Ap
c 20	89.5	5.1	29793	10 <td>US-09-973-451-38</td> <td>Sequence 38, Appli</td>	US-09-973-451-38	Sequence 38, Appli
21	87	4.9	5342	10 <td>US-09-852-922-1</td> <td>Sequence 1, Appli</td>	US-09-852-922-1	Sequence 1, Appli
22	86.5	4.9	1575	10 <td>US-09-815-242-7460</td> <td>Sequence 7460, Ap</td>	US-09-815-242-7460	Sequence 7460, Ap
c 23	86	4.9	890	10 <td>US-09-070-927A-606</td> <td>Sequence 606, App</td>	US-09-070-927A-606	Sequence 606, App
24	86	4.9	2457	9 <td>US-09-954-531-177</td> <td>Sequence 177, App</td>	US-09-954-531-177	Sequence 177, App
25	86	4.9	2457	9 <td>US-09-954-531-593</td> <td>Sequence 593, App</td>	US-09-954-531-593	Sequence 593, App
26	86	4.9	2457	10 <td>US-09-919-497-42</td> <td>Sequence 42, Appl</td>	US-09-919-497-42	Sequence 42, Appl
27	84	4.7	9783	10 <td>US-09-728-952-47</td> <td>Sequence 47, Appl</td>	US-09-728-952-47	Sequence 47, Appl
28	83	4.7	720	10 <td>US-09-910-943-733</td> <td>Sequence 733, App</td>	US-09-910-943-733	Sequence 733, App
29	82.5	4.7	2023	10 <td>US-09-765-272-199</td> <td>Sequence 199, App</td>	US-09-765-272-199	Sequence 199, App
30	82.5	4.7	368004	10 <td>US-09-949-654-3</td> <td>Sequence 3, Appli</td>	US-09-949-654-3	Sequence 3, Appli
31	82	4.6	2412	10 <td>US-09-801-368-393</td> <td>Sequence 393, App</td>	US-09-801-368-393	Sequence 393, App
32	80	4.5	1658	9 <td>US-09-890-813-15</td> <td>Sequence 15, Appli</td>	US-09-890-813-15	Sequence 15, Appli
33	80	4.5	2598	10 <td>US-09-815-242-7267</td> <td>Sequence 7267, Ap</td>	US-09-815-242-7267	Sequence 7267, Ap
34	80	4.5	3320	12 <td>US-10-047-757-1</td> <td>Sequence 1, Appli</td>	US-10-047-757-1	Sequence 1, Appli
35	79.5	4.5	1418	10 <td>US-09-974-300-7280</td> <td>Sequence 7280, Ap</td>	US-09-974-300-7280	Sequence 7280, Ap
c 36	79.5	4.5	2939	12 <td>US-10-044-090-350</td> <td>Sequence 350, App</td>	US-10-044-090-350	Sequence 350, App
37	79	4.5	1914	10 <td>US-09-815-242-4167</td> <td>Sequence 4167, Ap</td>	US-09-815-242-4167	Sequence 4167, Ap
38	79	4.5	1929	10 <td>US-09-815-242-8045</td> <td>Sequence 8045, Ap</td>	US-09-815-242-8045	Sequence 8045, Ap
39	78.5	4.4	1195	10 <td>US-09-969-708-594</td> <td>Sequence 594, App</td>	US-09-969-708-594	Sequence 594, App
40	78.5	4.4	1195	10 <td>US-09-880-107-3746</td> <td>Sequence 3746, Ap</td>	US-09-880-107-3746	Sequence 3746, Ap
41	78.5	4.4	2424	9 <td>US-10-041-007-14</td> <td>Sequence 14, Appli</td>	US-10-041-007-14	Sequence 14, Appli
42	78.5	4.4	2424	10 <td>US-09-887-586A-45</td> <td>Sequence 45, Appl</td>	US-09-887-586A-45	Sequence 45, Appl
43	78.5	4.4	2424	10 <td>US-09-903-012-45</td> <td>Sequence 45, Appl</td>	US-09-903-012-45	Sequence 45, Appl
44	78.5	4.4	2528	9 <td>US-10-041-007-13</td> <td>Sequence 13, Appl</td>	US-10-041-007-13	Sequence 13, Appl
45	78.5	4.4	3731	10 <td>US-09-822-268A-1</td> <td>Sequence 1, Appli</td>	US-09-822-268A-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-09-848-852A-4  
; Sequence 4, Application US/09848852A  
; Patent No. US20020106373A1  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; Tang, Y. Tom  
; Corley, Neil C.  
; Guegler, Karl J.  
; Yue, Henry  
; Patterson, Chandra  
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH EGF SIGNALING  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/848,852A  
; FILING DATE: 07-May-2001

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/069,725
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0515 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-845-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1991 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BEPNOT01
; CLONE: 2056178
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-848-852A-4

Alignment Scores:
Pred. No.: 3,72e-221 Length: 1991
Score: 1772.00 Matches: 332
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-848-852A-3 (1-332) x US-09-848-852A-4 (1-1991)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAlaHisLeuAla 20
DB 542 ATGTACTCTGCCAACCGTTTGGGGACAACTCTGGGCCAGAGGGGCGCATTGGCA 601
QY 21 ProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGln 40
DB 602 CCCCTGATGCCAGCATCCTCATCAGCAATGTGTGCAGCATCGGGGACCATGTGCCCG 661
QY 41 GluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLys 60
DB 662 GAGCTTTTTCAGGGCTCAGATTTGGGCTGCGCAGAGAGGCGCCTGGGGAGAAA 721
QY 61 AlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAsp 80
DB 722 GCGGCCAGCACAGCCCCCTCGGAGAGAGCATGTGTACCTCGCTACAGACATCTTGGAC 781
QY 81 GluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLys 100
DB 782 GAATTCCTTCAACGGTATGGCAGCTCATACCCCTCAGCATGATGAGGTAGTAGAAG 841
QY 101 LeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGln 120
DB 842 CTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAGGCGCTGGTTGCAG 901
QY 121 LeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAla 140
DB 902 CTGATCCAGTCTTACCAAGCGATGCCAGGCAATGCCATGTGTGAGGGCTTCCAGAGGCT 961
QY 141 TyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeu 160
DB 962 TATAAGCGGCAGCTGCTGACCATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCTC 1021
QY 161 AsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVal 180
DB 1022 AATGACAGGTGTGAACATGTATGGAGAGACCTGTGTATGGACAGAGTCCCTGAAAAGGTG 1081
QY 181 HisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLys 200
DB 1082 CATTTCTCAATAGTTTCTTATGATAAAATCCGTACCAAGGTTATGATGGGGTAAA 1141
QY 201 ArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuIleProIleHisLeu 220
```

```
DB 1142 AGGTGGACCAAAAACGTGGACATCTTCAATAAGGAGCTACTGCTTAATCCCCATCCACTG 1201
QY 221 GluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAsp 240
DB 1202 GAGTGATGGTCCCTCATCTCTGTTGATGTGAGGGGACGACCATCACCATTATTTGAC 1261
QY 241 SerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGlu 260
DB 1262 TCGCAGCGGTACCTTAACCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAGAG 1321
QY 261 AlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAsn 280
DB 1322 GCGGTAAGAAAGACCGACTGGATTCCACCAGGCGCTGGAAGGTTACTTCAAAATGAAT 1381
QY 281 ValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHis 300
DB 1382 GTGCCAGGAGCAATAATGACGTGCTGTTGTTGTTGTCAGTACTGCAAGCAT 1441
QY 301 LeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgGln 320
DB 1442 CTGGCCCTGTCTAGCCATTTCAGCTTCACCAGCAGGACATGCCCAAACTTCGTCGCGAG 1501
QY 321 IleTyrLysGluLeuCysHisCysLysLeuThrVal 332
DB 1502 ATCTACAAGGAGCTGTGTCACTGCAAACTCACGTGTG 1537

RESULT 2
US-09-848-852A-3 (1-332) x US-09-848-852A-4 (1-1991)
; Sequence 835, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 835
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 29-LIB188-017-Q1-E1-H1
US-09-848-852A-3 (1-332) x US-09-848-852A-4 (1-1991)

Alignment Scores:
Pred. No.: 8.19e-83 Length: 432
Score: 704.00 Matches: 135
Percent Similarity: 97.86% Conservative: 2
Best Local Similarity: 96.43% Mismatches: 3
Query Match: 39.73% Indels: 0
DB: 10 Gaps: 0

US-09-848-852A-3 (1-332) x US-09-848-852A-4 (1-1991)

QY 28 IleSerAsnValCysSerIleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAsp 47
DB 431 ATCAGCAACGTGTGCAGCATCGGGGACCCAGCTGCCAGGAGCTTTTTCAGGGCTCAGAT 372
QY 48 LeuGlyMetAlaGluGluAlaGluArgProGlyGluLysAlaGlyGlnHisSerProLeu 67
DB 371 GTGGGCACTCAGAAAGAGGCTGAGAGGCTGGGGAGAAAGCTGGCCAGCAGCCCTCTG 312
QY 68 ArgGluGluHisValThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGly 87
```

```

|||||
Db 311 CGGAGGAGCATGTACCTGTGTGAGAGCATCTTGGATGAATTCCTTCAGACTTACGGC 252
QY 88 SerLeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGln 107
Db 251 AGCCTCATACCCCTCAGACATGATGAGTAGTGGAGAACTAGAGGACATTTCCAGAG 192
QY 108 GluPheSerThrProSerArgLysGlyLeuValValLeuGlnLeuIleGlnSerTyrGlnArg 127
Db 191 GAGTTCCTGCACCTTCCAGGAAGGCCCTGGTGTCTGCAGTTGATCCAGTCATACACGG 132
QY 128 MetProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThr 147
Db 131 ATGCCAGGCAATGCCATGTGTGAGGGCTTCCGAGTCACTATAAGCGCATGTGTCACC 72
QY 148 MetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMet 167
Db 71 ATGGATGACCTGGGAACCTTTGTATGGACAACTGGCTCAATGACCAAGGTGTGAACATG 12
RESULT 3
US-09-777-564-1710/c
; Sequence 1710, Application US/09777564
; Patent No. US2002022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777.564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 1710
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-777-564-1710

```

```

Alignment Scores:
Pred. No.: 3,7e-77 Length: 503
Score: 661.50 Matches: 125
Percent Similarity: 88.03% Conservative: 0
Best Local Similarity: 88.03% Mismatches: 0
Query Match: 37.33% Indels: 17
Db: 10 Gaps: 1

US-09-848-852A-3 (1-332) x US-09-777-564-1710 (1-503)
QY 191 LeuArgThrLysGlyTyrAspGlyValLysArgTyrTrpThrLysAsnValAspIlePheAsn 210
Db 502 CTCCTACCAAGGTTATGATGGGTGAAAGGTGGACCAAAACGTGGACATCTTCAT 443
QY 211 LysGluLeuLeuLeuIleProIleHisLeuGluValHisTyrSerLeuIleSerValAsp 230
Db 442 AAGGAGCTACTGCTAATCCCATCCACCTCGAGGTGCTATGCTCCTCTCTGAT 383
QY 231 ValArgArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgCysPro 250
Db 382 GTGAGCGGACGCACCATCACCTATTATTGACTCGACGCTACCTAAACCGCGCTGCCCT 323
QY 251 LysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHis 270
Db 322 AGCATATTGCCAGTACTACAGCAGCGGCGGTAAAGAAACCGACTGGATTTCCAC 263
QY 271 GlnGlyTyrLysGlyTyrPheLysMetAsnValAlaArgGlnAsnAspSerAspCys 290
Db 262 CAGGCTGGAAAGGTTACTTCAAAATG----- 236
QY 291 GlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThr 310
Db 235 -----TACTGCAAGCATCTGGCCCTGCTCAGCCCATTCAGCTTCACC 194

```

```

QY 311 GlnGlnAspMetProLysLeuArgGlnIleTyrLysGluLeuCysHisCysLysLeu 330
Db 193 CAGCAGGACATGCCCAACTTCGTGGCAGATCTACAAGGAGCTGTGTCACTGCAACATC 134
QY 331 ThrVal 332
Db 133 ACTGTG 128
RESULT 4
US-09-856-247A-1
; Sequence 1, Application US/09856247A
; Patent No. US20020151028A1
; GENERAL INFORMATION:
; APPLICANT: Lima, Christopher
; APPLICANT: Mossesova, Elena
; TITLE OF INVENTION: Structure-Based Drug Design for Ulpl Protease Substrates
; FILE REFERENCE: 2650/LG681-US1
; CURRENT APPLICATION NUMBER: US/09/856.247A
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 60/205,336
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-856-247A-1
Alignment Scores:
Pred. No.: 6,21e-19 Length: 1866
Score: 222.50 Matches: 77
Percent Similarity: 43.46% Conservative: 56
Best Local Similarity: 25.16% Mismatches: 120
Query Match: 12.56% Indels: 53
Db: 10 Gaps: 12

US-09-848-852A-3 (1-332) x US-09-856-247A-1 (1-1866)
QY 36 AspHisValAlaGlnGluLeu---PheGlnGlySerAspLeuGlyMetAlaGluAla 54
Db 1063 GATTACTTAAATCAAAAGTTGAAGTTTGATAGTATATATTAGAGTTTGAGAAAGACTTTC 1122
QY 55 GluArgProGlyGluLysAlaGlyGlnHisSerProLeuArgGluGluHisValThrCys 74
Db 1123 AAAAGATATACGAATTTTAAATCAAAAGAGAGATTCAGAAGATCTTAAAAAAG 1182
QY 75 ValGlnSerIleLeuAspGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThr 94
Db 1183 AAAGAACAATTCGCCCAAGAG-----AAACTGTGTCCT----- 1215
QY 95 AspGluValValGluLysLeuGluAspIlePheGlnGlnGluPheSerThrProSerArg 114
Db 1216 ---GAATTAATGAAAGACGATGCCAAGTACAAAAGCTTTGGCA-----TCTAGA 1266
QY 115 LysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetVal 134
Db 1267 GAAATACCTCACTTAATGATAGATATATAGAG----- 1302
QY 135 ArgGlyPheArgValAlaTyrLysArgHisValLeuThrMetAspLeuGlyThrLeu 154
Db 1303 -----ATAACAGTACGTGATTTTAAAGACCTTG 1329
QY 155 TyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAsp 174
Db 1330 GCACCACGAAAGTGCCTAAATGACACTATCATTTAGTTTATGAAATFACATTTGAAAAA 1389
QY 175 ThrValProGluLysValHisPheAsnSerPhePheTyrAspLysLeuArgThrLys 194
Db 1390 TCTACCCCTTAATACAGTG---CCGTTTATTCATTTTCTATACCAATTTATCAAGAGG 1446
QY 195 GlyTyrAspGlyValLysArgTrpThrLys-----AsnValAspIlePheAsn 210
Db 195 ||||| |||||:||||| |||

```

```
Db 1447 GGTATCAAGCGTCGGAGGTGGATGAAGAGAAAGACACAAATTGATAAATTTGAT 1506
QY 211 LysGluLeuLeuLeuLeuProIleHisLeu---GluValHisTrpSerLeuLeuSerVal 229
Db 1507 AAA-----ATCTTTACACCAATAAAATTTGAACCAATCCCACACTGGCGTTGGGCATAAAT 1560
QY 230 AspValArgArgThrIleThrThrPheAspSer-----GlnArgThrLeu 245
Db 1561 GATTAAAGAAAGAAACTATAGGTTAGCTAGATTCAATTATCGAATGGTCCCAATGCTATG 1620
QY 246 AsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAsp 265
Db 1621 AGTTTCCTACTACTGACTGACTGCTGCAAAATATGTTATGGAGGAAGTAAGCATACAATA 1680
QY 266 ArgLeuAspPheHisGlnGlyTyrPheLysGlyTyrPheLysMetAsnValAlaArgGlnAsn 285
Db 1681 GGAGAACACTTT-----GATTTGATTCATTATAGATTCTCTATGGAAGTGCAGATGCG 1725
QY 286 AsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerGln 305
Db 1726 AATGGCTACGACTGCTGCAATATATGTTGTATGAATACTCTCTATGGAAGTGCAGATGCG 1785
QY 306 ProPheSerPheThrGlnGlnAspMetProLysLeuArgArgGlnIleTyrLysGluLeu 325
Db 1786 CCATTGGATTGATTATATAAGATGCGATTAGGATGAGAAGATTATT----- 1833
QY 326 CysHisCysLysLeuThr 331
Db 1834 GCCATTGATTTAACC 1851
RESULT 5
US-09-867-701-4735
; Sequence 4735, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Agiate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4735
; LENGTH: 435
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-4735
Alignment Scores:
Pred. No.: 4.67e-12 Length: 435
Score: 162.00 Matches: 36
Percent Similarity: 68.25% Conservatives: 7
Best Local Similarity: 57.14% Mismatches: 11
Query Match: 9.14% Indels: 10
DB: 10 Gaps: 2
US-09-848-852A-3 (1-332) x US-09-867-701-4735 (1-435)
QY 269 PheHisGlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAsnAspSer 288
Db 255 TTCTCTGAAGATGGAGA-----CACATCTCATATGAAATGTGTAGCACAGTCTCT 305
QY 289 AspCysGlyAlaPheValLeu-----GlnTyrCysLysHisLeuAla 302
Db 306 GACAC-GGGGGGTTTCTCATGGCTTCTGTTTAAACACCCAGTACTGCAAGCATCTGGCC 364
QY 303 LeuSerGlnProPheThrGlnGlnAspMetProLysLeuArgArgGlnIleTyr 322
Db 365 CTCTCTCAGCCATTAGCTTACCCAGAGGACATGCCCAAACTTCTGTCGGCAGATCTAC 424
```

```
QY 323 LysGluLeu 325
Db 425 AAGGAGCTG 433
RESULT 6
US-09-960-253-74
; Sequence 74, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 5540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-960-253-74
Alignment Scores:
Pred. No.: 6.35e-08 Length: 5540
Score: 144.00 Matches: 34
Percent Similarity: 53.70% Conservatives: 24
Best Local Similarity: 31.48% Mismatches: 29
Query Match: 8.13% Indels: 21
DB: 10 Gaps: 3
US-09-848-852A-3 (1-332) x US-09-960-253-74 (1-5540)
QY 143 ArgHisVal-LeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAs 162
Db 1992 AGGCATCTCTGTACCAATGAGGACCTGCACCTCTCTAAATGAAGGAGAATTTTAAATGA 2051
QY 162 pGlnValMetAsnMetTyr---GlyAspLeuValMetAspThrVal-----Pr 177
Db 2052 TGTATTATAGACTTTTATTTGAAATACITGGTGCTTGAAAAAAGCGCATGGAAGAGACGC 2111
QY 177 oGluLysValHisPheAsnSerPhePheTyrAspLysLeu----- 191
Db 2112 TGACCGAATTCATATATTCAGTCTTTTCTATAACCCCTTAATCAGACAGAGGAG 2171
QY 192 -----ArgThrLysGlyTyrAspGlyValLysArgTr 202
Db 2172 AAATCATGAAACAACCTAATCTGTCTAATACAGCAAAAACGCGATGGGAGAGTAAAAACATG 2231
QY 202 pThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGluVa 222
Db 2232 GACCCGGCAGTAGATATTTTGAGAAGGATTATTATTTTGTACCCCTTAATGAAGCTGC 2291
QY 222 lHisTrpSerLeuIleSerVal 229
Db 2292 ACACGTGTTTTTGGCTGTTGTT 2313
RESULT 7
US-09-878-574-7902
; Sequence 7902, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
```



```
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 7902
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701100484H1
US-09-878-574-7902

Alignment Scores:
Pred. No.: 2,7e-08 Length: 264
Score: 130.50 Matches: 27
Percent Similarity: 52.50% Conservative: 15
Best Local Similarity: 33.75% Mismatches: 25
Query Match: 7.36% Indels: 13
DB: 10 Gaps: 2

US-09-848-852A-3 (1-332) x US-09-878-574-7902 (1-264)
QY 178 GluLysValHisPheAsnSerPheTyrAspLysLeu----- 191
Db 22 AAGAGGTTTCACCTTTTCAATTGCTTTTCTTCTTAAGCTTGTGATTAAAGCACAGAT 81
QY 192 ---ArgThrLysGlyTyrAspGly-----ValLysArgTrpThrLys 204
Db 82 AATCGTTCAATTGCTTGTGATGTAAGCAGCATTTTCAGCGTGTAAAGCAATTGGACAAGA 141
QY 205 AsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHisTrp 224
Db 142 AAGTGAACCTTTTGAAGAGGATTATCTTCATTCCCATAAACTATAGTCTTCATTGG 201
QY 225 SerLeuIleSerValAspValArgArgGTrpIleThrTyrPheAspSerGlnArgTrp 244
Db 202 AGTTGATGTCATTTGTCACCTGCTGAAGTCATGATGTCATGATGATGATGATGATGAT 261

RESULT 8
US-10-008-461-1
; Sequence 1, Application US/10008461
; Patent No. US20020127692A1
; GENERAL INFORMATION:
; APPLICANT: Barbara Ink
; APPLICANT: Alan Lewis
; TITLE OF INVENTION: New Protein
; FILE REFERENCE: QG1034(p80209)
; CURRENT APPLICATION NUMBER: US/10/008,461
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 0027905.9
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 639
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(636)
US-10-008-461-1

Alignment Scores:
Pred. No.: 1.25e-06 Length: 639
Score: 122.50 Matches: 48
Percent Similarity: 40.40% Conservative: 32
Best Local Similarity: 24.24% Mismatches: 73
Query Match: 6.91% Indels: 45
DB: 12 Gaps: 8

US-09-848-852A-3 (1-332) x US-10-008-461-1 (1-639)

QY 139 ValAlaTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlnAsn 158
Db 16 TTGAGTTATGACAGCTACTGCTGGCAATCAGATGTCCTACTATTGGATCGGCAAGC 75

; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 7902
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701100484H1
US-09-878-574-7902

Alignment Scores:
Pred. No.: 2,7e-08 Length: 264
Score: 130.50 Matches: 27
Percent Similarity: 52.50% Conservative: 15
Best Local Similarity: 33.75% Mismatches: 25
Query Match: 7.36% Indels: 13
DB: 10 Gaps: 2

US-09-848-852A-3 (1-332) x US-09-878-574-7902 (1-264)
QY 178 GluLysValHisPheAsnSerPheTyrAspLysLeu----- 191
Db 22 AAGAGGTTTCACCTTTTCAATTGCTTTTCTTCTTAAGCTTGTGATTAAAGCACAGAT 81
QY 192 ---ArgThrLysGlyTyrAspGly-----ValLysArgTrpThrLys 204
Db 82 AATCGTTCAATTGCTTGTGATGTAAGCAGCATTTTCAGCGTGTAAAGCAATTGGACAAGA 141
QY 205 AsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHisTrp 224
Db 142 AAGTGAACCTTTTGAAGAGGATTATCTTCATTCCCATAAACTATAGTCTTCATTGG 201
QY 225 SerLeuIleSerValAspValArgArgGTrpIleThrTyrPheAspSerGlnArgTrp 244
Db 202 AGTTGATGTCATTTGTCACCTGCTGAAGTCATGATGTCATGATGATGATGATGATGAT 261

RESULT 8
US-09-728-445-223
; Sequence 223, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
; FILE REFERENCE: LEX-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 223
; LENGTH: 100
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-728-445-223

Alignment Scores:
Pred. No.: 5.11e-07 Length: 100
Score: 115.50 Matches: 22
Percent Similarity: 84.85% Conservative: 6
Best Local Similarity: 66.67% Mismatches: 4
Query Match: 6.52% Indels: 1
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x US-09-728-445-223 (1-100)

QY 180 ValHisPheAsnSerPheTyrAspLysLeuArgThrLysGlyTyrAspGlyVal 199
Db 1 GTCCACTTCTTCAACAGCTTTTCCATAGACAGCTGGTAACCAAGGCTATAATGGAGTT 60
```

```
QY 159 TrpLeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGlu 178
Db 76 TGSCCTCAATGACCATATTATTGGTTTGGTTTGAG----- 111
QY 179 LysValHisPheAsnSerPheTyrAsp----- 189
Db 112 -----TACTTTGCCAACAGTCAGTTTCATGACTGCTCTGATCAGCTCAGTTTCATCAGC 165
QY 190 -----LysLeuArgThrLysGlyTyrAspGlyValLysArgTrp 202
Db 166 CTTGAAGTCACCCAGTTTCATCAAGTC-----ACTAGCAACCCAGCAGATTGCCATGTTCC 222
QY 203 ThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGlu--- 221
Db 223 CTTGAACCACTGGACCTCCCAACAGAGAGTTGTTATTTTAGCCATCAATGATAACTCC 282
QY 222 -----ValHisTrpSerLeuIleSerValAspValArgArgTrp 235
Db 283 AACAGGCGAGCTGGAGGAACCCACTGGAGTTTATTGCTCTACCTCCCAAGATAAAATAGC 342
QY 236 IleThrTyrPheAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLys 255
Db 343 TTTTTCATTATGATTTCCCATAGCAGGAGCACTCAGTTCCACCAAGCAGGAGTAGCAGAG 402
QY 256 TyrLeuGlnAlaGluAlaValLysLys---AspArgLeuAspPheHisGlnGlyTrpLys 274
Db 403 AAACCTGGAGGCTTCTTAGGCAGAAAAGGAGACAAACTGGCCTTT----- 447
QY 275 GlyTyrPheLysMetAsnValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheVal 294
Db 448 -----GTGGAAGAGAGAAGCCCTGCCCCAACAAACAGCCTATGACTGTGGATGTACGTG 501
QY 295 LeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGln 312
Db 502 ATA-----TGTAACACTGAGGCTTGTGTGACAACTTC---TTTAGGCAACAG 546

RESULT 9
US-09-728-445-223
; Sequence 223, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
; FILE REFERENCE: LEX-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 223
; LENGTH: 100
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-728-445-223

Alignment Scores:
Pred. No.: 5.11e-07 Length: 100
Score: 115.50 Matches: 22
Percent Similarity: 84.85% Conservative: 6
Best Local Similarity: 66.67% Mismatches: 4
Query Match: 6.52% Indels: 1
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x US-09-728-445-223 (1-100)

QY 180 ValHisPheAsnSerPheTyrAspLysLeuArgThrLysGlyTyrAspGlyVal 199
Db 1 GTCCACTTCTTCAACAGCTTTTCCATAGACAGCTGGTAACCAAGGCTATAATGGAGTT 60
```







Db Qy Db Qy Db Qy Db Qy Db Qy Db Qy Db Qy Db Qy Db Qy Db Qy Db Qy Db Qy Db Qy Db Qy Db Qy

463 CCAGATTCGTGTTATTCTTTTTCAACTTTTACCGAAT----- 498  
42 LeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLysAla 61  
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
499 TTGTATCCTGGTACACCTTTAGCAACT-----GATATAGTT 534  
62 GlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAspGlu 81  
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
535 GGAAGTGAGGAGTCATTTCCCAAACAATCTAACTAAATTTGCAGAATAATTTACAAAG 594  
82 PheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluValValGluLysLeu 101  
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
595 TTTTACAAACCTGTAACATGCTCTCGTTTTTTAGTTGGTAAATTTTCATGTGGACCGAGTA 654  
102 GluAspIlePheGln-----GlnGluPheSerThrProSerArgLys 115  
:|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
655 CAGGACTATTTTGAAGCAAAAGAACGAAAGATTCAGATTTTCAGGAAGTAGCAAGAGAA 714  
116 GlyLeuValLeuGlnLeuIleGlnSerTyrglnArgMet-----Pro 129  
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
715 AAGTGTGTTTTACAGCCTGTAAAGGCCAACAGATAGTAGAGAAATGGAAGTATCTCTCCC 774  
130 GlyAsnAlaMet---ValArgGlyPheArgVal-----AlaTyrllysArg 143  
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
775 AAACACCGANTGGAGTTAGAGGTAGCGAGAGAGAGTTCTTCAACCGGATGCTATCGACAT 834  
144 HisValLeuThrMetAspaspLeuGlyThrLeutyrglyGlnAsnTrpLeuAsnAspGln 163  
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
835 CATATTTTATAAAAATTATTGTTTCCAATCATCTTTGGT-----TGGACTTCGGATCGT 888  
164 ValMetAsnMetTyrglyAspLeuValMetAspThr-----Val 176  
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
889 TTTCAAAAAATGTTATGAATCAGGTAAATGATGCATPCCTTATCTCTGGAAGTTGAAGTA 948  
177 ProGluLysValHisPhePheAsnSerPhePhetyrAspLys-----LeuArg 192  
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
949 ACAAGTCGCTTCATTTTGTCATGTCACAATGGATACGAAAGCCAGTTCGTTTGCT 1000  
193 ThrLysGlyTyraSpGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGlu 212  
::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1009 CATCAATTTAGGAAGGCTATTCGTAATTTTACAAAGGATTTAGATATTTACAGAGAACAT 1060  
213 LeuLeuLeuIleProIleHisLeu-----GluValHisTrpSerLeuIleSerValasp 230  
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1069 TTAGATATTATCAAAAGAGAGAGTGTGGCGGAATTTTTCAGTAGCATGAACCTCTGTGAA 112  
231 ValArgArgArgThrIleThrTyrglyPheAspSerGlnArgThrLeuAsnArgArgCysPro 250  
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1129 TTTTATTCACCGCAATATGATGCTTTTGAAATGGTGAGACAATT----- 117  
251 LysHisIleAlaLysTyrrLeuGlnAlaGluAlaValLysLysAspArgLeuAsp 268  
::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
1174 TTTGATTTGGCGAAAAATTTTACAG--GAAATTTACTTTTACAGAGATGCTCTTGAT 1224

Alignment Scores:	
Pred. No.:	0.0578
Score:	90.50
Percent Similarity:	29.21%
Best Local Similarity:	20.86%
Query Match:	5.11%
DB:	10
Length:	1384
Matches:	58
Conservative:	51
Mismatches:	11
Indels:	55
Gaps:	12

QY 22 ProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGlnGlu 41

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 19:36:15 ; Search time 257 seconds  
(without alignments)  
124.189 Million cell updates/sec

Title: US-09-848-852A-3

Perfect score: 1772

Sequence: 1 MYSAQREWGTTIARRGAHLA.....DMPKLRQIYKELCHCKLTV 332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1667	94.1	371	2 T08759	conserved hypothe
2	319.5	18.0	568	2 T39840	conserved hypothe
3	268.5	15.2	662	2 T16845	hypotheical prote
4	240.5	13.6	233	2 T01233	hypotheical prote
5	222.5	12.6	621	2 S63462	hypotheical prote
6	205.5	11.6	652	2 T37822	probable centromer
7	197.5	11.1	424	2 H85175	hypotheical prote
8	196	11.1	1034	2 S49947	SMT4 protein - yea
9	193	10.9	547	2 T02274	hypotheical prote
10	190	10.7	398	2 E71424	hypotheical prote
11	163	9.2	710	2 T04979	hypotheical prote
12	156	8.8	1240	2 T48800	SMT4 related prote
13	136	7.7	920	2 T26650	hypotheical prote
14	124.5	7.0	342	2 T19877	hypotheical prote
15	118.5	6.7	311	2 T32915	hypotheical prote
16	116.5	6.6	662	2 T33900	hypotheical prote
17	109.5	6.2	1121	2 T13796	hypotheical prote
18	109	6.2	1305	2 H96559	kinesin-related pr
19	107.5	6.1	272	2 C86642	hypotheical prote
20	106	6.0	418	2 D71460	probable membrane
21	105	5.9	1153	2 F84468	hypotheical prote
22	105	5.9	1272	2 C64513	hypotheical prote
23	104.5	5.9	272	2 C86884	transposase of IS1
24	103	5.8	1218	2 E84537	hypotheical prote
25	102	5.8	550	2 G84500	hypotheical prote
26	101.5	5.7	780	2 B84084	ATP-dependent DNA
27	99.5	5.6	1121	2 T13750	kinesin-like prote
28	98	5.5	227	2 D86631	transposase of IS1
29	98	5.5	1314	2 T47331	hypotheical prote

30	97	5.5	625	2 G72420	oligopeptide ABC t
31	97	5.5	790	1 WMFM9	90K protein - alfa
32	96.5	5.4	272	2 D86671	transposase of IS1
33	95.5	5.4	543	2 T02022	hypotheical prote
34	95.5	5.4	1201	2 F86386	hypotheical prote
35	94.5	5.3	929	2 B81982	probable isoleucin
36	94	5.3	227	2 G86704	transposase of IS1
37	93.5	5.3	662	2 E97738	DNA topoisomerase
38	93.5	5.3	1472	2 A84470	hypotheical prote
39	93	5.2	445	2 F83881	hypotheical prote
40	92.5	5.2	273	2 A42549	hypotheical prote
41	92	5.2	1198	2 E86402	hypotheical prote
42	91.5	5.2	353	2 D81273	probable 3-oxoacyl
43	91.5	5.2	661	2 T30179	hypotheical prote
44	91	5.1	1175	2 D85089	hypotheical prote
45	90.5	5.1	297	2 C70231	hypotheical prote

#### ALIGNMENTS

##### RESULT 1

T08759

conserved hypotheical protein DKF2p586K0919.1 - human

C:Species: Homo sapiens (man)

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 13-Aug-1999

C:Accession: T08759

R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z16471

A:Accession: T08759

A:Molecule type: mRNA

A:Residues: 1-371 <WAM>

A:Cross-references: EMBL:AL050283

A:Experimental source: adult uterus; clone DKF2p586K0919

C:Genetics:

A:Note: DKF2p586K0919.1

Query Match

Best Local Similarity 94.1%; Score 1667; DB 2; Length 371;

Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	19	LAPPDASILISNVCISGIDHVAQELFGQSDLGMAEEAERPEKAGQHSPLREHVTVCVSI	78
Db	58	LAPPDASILISNVCISGIDHVAQELFGQSDLGMAEEAERPEKAGQHSPLREHVTVCVSI	117
Qy	79	LDEFLTQYTGSLIPLSTDEVVEKLEDFQOEFTSPSRKGLVLQLIQSYQRMPCNVMVRGPR	138
Db	118	LDEFLTQYTGSLIPLSTDEVVEKLEDFQOEFTSPSRKGLVLQLIQSYQRMPCNVMVRGPR	177
Qy	139	VAYKRHVLTMDDLGTLYGONLNDQVMNMGDLVMDTVPEKVHFFNSFFYDKLRTKGYDG	198
Db	178	VAYKRHVLTMDDLGTLYGONLNDQVMNMGDLVMDTVPEKVHFFNSFFYDKLRTKGYDG	237
Qy	199	VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRCCKPHIAKYIQ	258
Db	238	VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRCCKPHIAKYIQ	297
Qy	259	AEAVKRDRLDFHQWKGFKMNVARQNNDSCGAFVLQYCKHLALSQPFSTQDMPKLR	318
Db	298	AEAVKRDRLDFHQWKGFKMNVARQNNDSCGAFVLQYCKHLALSQPFSTQDMPKLR	357
Qy	319	ROIYKELCHCKLTV	332
Db	358	ROIYKELCHCKLTV	371

##### RESULT 2

T39840

conserved hypotheical protein SPBC19G7.09 - fission yeast (Schizosaccharomyces pombe

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T39840

R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.  
submitted to the EMBL Data Library, February 1998  
A:Reference number: Z21884  
A:Accession: T39840  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-568 <LN>  
A:Cross-references: EMBL:AL021839; PIDN:CAA17063.1; GSPDB:GNO0067; SPDB:SPBC19G7.09  
A:Experimental source: strain 972h-; cosmid c19G7  
C:Genetics:  
A:Gene: SPDB:SPBC19G7.09  
A:Map position: 2

Query Match 18.0%; Score 319.5; DB 2; Length 568;  
Best Local Similarity 28.7%; Pred. No. 1.6e-18;  
Matches 90; Conservative 51; Mismatches 114; Indels 59; Gaps 9;  
QY 59 EKAGQHSLPREEHVTCVQS-----ILDEFLQYGSILPLSTDEVEKLEDFIQ----- 107  
Db 264 KSPHDSTSRKASRFVQSDQPARNIIVTSDIQNEKSLLLIIRD-LKEKQTESFQDQWNEV 322  
QY 108 -----EFSTPSRKGLVLQLIQSYQRMPCGNMAYRGFRVAYKRH 144  
Db 323 DFQLKGLSTPPTRPKFPELEFPDNARKALKYLQNSVSSSPILITKENIP----- 378  
QY 145 VLTMDLGLTYGQNLNDQVMNMYGDLV-----MDTVPEKVVHFFNSFFYDKLRTKGYDGV 199  
Db 379 -ITLKDLHLTRNRQWLNDVFNFMNLISERSKIDSLPRVHGNTFFYTSLQRRGAGV 437  
QY 200 KRWTK--NVDIFNKELLIPHLEVHWSLISVDVRRRTTYFDSQRTLNRRCPKHIAYL 257  
Db 438 RRAKARVNIADMDAVFIPVHLVDHWCMAVINKSKRFEYWDLAGSPGKVFLLRDY 497  
QY 258 QAEAVKKDRDLDFHQGWGKFKMVARQNDSCGAFVLQYCKHLALSQPFSTQDMPKL 317  
Db 498 IAET--KGAVDV--SDWENFMDNSPQRNGHDCGVFACTAECVSRVNPVQFSQNDMP 554  
QY 318 R-----RQIV 322  
Db 555 RIKMAASIIDAQIV 568

RESULT 3  
T16845  
hypothetical protein T10F2.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T16845  
submitted to the EMBL Data Library, March 1995  
A:Description: The sequence of C. elegans cosmid T10F2.  
A:Reference number: Z18589  
A:Accession: T16845  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-662 <MIL>  
A:Cross-references: EMBL:U23412; NID:g727446; PID:g727449; PIDN:AAA64293.1; CESP:T10F2.3  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:T10F2.3  
A:Introns: 17/1; 51/1; 96/3; 197/3; 291/3; 342/3; 375/2; 619/2; 645/3; 649/1

Query Match 15.2%; Score 268.5; DB 2; Length 662;  
Best Local Similarity 26.4%; Pred. No. 3.3e-14;  
Matches 75; Conservative 53; Mismatches 79; Indels 77; Gaps 10;  
QY 27 LISNVCISG---DHVAQELFQGSDLGMAEAPRGKAGQHSPLREEHVTCVQSILDEFL 83  
Db 396 ILQINSIGLSGAYSARGPQRYONS---YQLSKQKEDKLLLEEARIREGRHS----- 441  
QY 84 QTYGSLIPLSTDEVEKLEDFIQEESTPSRKGLVLQLIQSYQRM----- 128  
Db 442 QTRG-----DRLEDV-----RKRLQGIATIRPKVEKKKVDPMALPDAA 481

QY 129 -----PGNAMVGRFVAYKRHVLTMDLGLTYGQNLNDQVMNMYGDLVM--- 173  
Db 482 DALVERAWSGNPNQFVDAFSIQ-----ICKKDLATLSGLHWNDEIINFYLQLICDRS 536  
QY 174 --DTVPEKVVHFFNSFFYDKLRTKGYGVRKWTAKNVDIFNKELLIPHLEVHWSLISVDV 231  
Db 537 NGDSKYPKIYAFNTFFYSNIVSGYASVAKRTRKVDIFAFDVLVPLVHGLHWCMAVIDM 596  
QY 232 RRTTYFDSQRTLNRRCPKHIAYLQAEAV--KKDRLDPHQGW 273  
Db 597 GERKIEFYDSLDYDNTAVLPLALRGYLEAESLDEKDKKKTAMNF-SGW 639

RESULT 4  
T01233  
hypothetical protein F6N23.7 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 22-Oct-1999  
C:Accession: T01233  
R:Geisel, C.  
submitted to the EMBL Data Library, April 1998  
A:Description: The sequence of A. thaliana F6N23.  
A:Reference number: Z14281  
A:Accession: T01233  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-233 <GEI>  
A:Cross-references: EMBL:AF058919; NID:g3047100; PID:g3047118; GSPDB:GNO0063; ATSP:F6  
C:Genetics:  
A:Gene: ATSP:F6N23.7  
A:Map position: 5  
A:Introns: 9/1; 21/2; 52/3; 85/3; 115/3; 147/3; 182/2; 213/3

Query Match 13.6%; Score 240.5; DB 2; Length 233;  
Best Local Similarity 31.3%; Pred. No. 1.7e-12;  
Matches 62; Conservative 32; Mismatches 75; Indels 29; Gaps 7;  
QY 149 DDLGLTYGQNLNDQVMNMYGDLVMDTV---PE---KVHFFNSFFYDKLRT---KGYDGV 199  
Db 38 ETQLQCLRPQWLNDVTLNLYLELKERQTRDPQKYFKCHFFNTFFYVKLVSGSGYNYKAV 97  
QY 200 KRWTK----NVDIFNKELLIPHLEVHWSLISVDVRRRTTYFDSQRTLNRRCPKHIAK 255  
Db 98 SWTTTKRKLGYDLDCDIIFVPIHIDIHTLGTGINNRKRFVYLDLSLT-----GAK 149  
QY 256 YLQAEAVKKDRDLDFH--QGWKGYFKMVARQNDSCGAFVLQYCKHLALSQPFSTQ---- 311  
Db 150 YLVDEVKQSKQKNIDVSSWGMVEYVEERPOQNGYDCGMFLKYIDFYSRGLSLQFSQVIR 209  
QY 312 ----QDMPKLRRQIYKEL 325  
Db 210 DVTKKDMPYFRJRTAKEI 227

RESULT 5  
S63462  
hypothetical protein YP020c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein LPB1lc  
C:Species: Saccharomyces cerevisiae  
C:Date: 16-May-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002  
C:Accession: S63462  
R:Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.;  
submitted to the EMBL Data Library, September 1995  
A:Reference number: S63452  
A:Accession: S63462  
A:Molecule type: DNA  
A:Residues: 1-621 <WAN>  
A:Cross-references: EMBL:U36624; NID:g1276642; PID:g1039457; GSPDB:GNO0016; MIPS:YP0  
C:Genetics:  
A:Gene: SGD:ULP1; MIPS:YP020c  
A:Cross-references: SGD:S0005941  
A:Map position: 16L





Best Local Similarity 27.2%; Pred. No. 6.2e-08;  
Matches 52; Conservative 33; Mismatches 64; Indels 42; Gaps 5;

QY 144 HVLTMDDLGLTYGONWLNDOVMNMYGDLVMDTVPEK-----VHFNSFFYDKLRTKG- 195

Db 442 YTTNQDFKCLFNKNDWNSIDLFTFKFYFESSIEKSIIRKQVHLMSSFFYTKLISNPA 501

QY 196 --YDGVKRWTKNVDIFNKELLIPILHLEHVH-----SLI 227

Db 502 DYYSNVKKWNTDLSFKYVVPINISYHWFSCIITNLDAILDHONKDKNDAINSDEI 561

QY 228 SDVVRRTITYFDSQRTLNRCRKPHIAKYLOAEAVKKDRDLDFHQGWGYFKMN---VARQ 284

Db 562 SINPLNVLITFDSLRQTHREIDPIKEFLISYALDKYSIQLD---KTQIKMKTKCPVPQ 618

QY 285 NNDSDCGAFVL 295

Db 619 PNMSDCGVHVI 629

RESULT 9  
T02274  
hypothetical protein T13D8.11 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 22-Oct-1999  
R:Vysotskaia, V.S.; Schwartz, J.R.; Kwan, A.; Toriumi, M.; Yu, G.; Oji, O.; Liu, S.; Li-  
rtz, D.; Li, Y.; Palm, C.J.; Shim, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N.  
submitted to the EMBL Data Library, June 1998  
A:Description: Arabidopsis thaliana chromosome 1 BAC T13D8 sequence.  
A:Reference number: Z14649  
A:Accession: T02274  
A:Molecule type: DNA  
A>Status: translated from GB/EMBL/DBJ  
A:Residues: 1-547 <Y>  
A:Cross-references: EMBL:AC004473; NID:g3108025; PID:g3249071; GSPDB:GN00059; ATSP:T13D8

A:Gene: ATSP:T13D8.11

A:Map position: 1

A:Introns: 106/1; 154/3; 259/2; 275/3; 304/2; 332/2; 366/3; 399/2; 440/2; 478/3; 514/3

Query Match 10.9%; Score 193; DB 2; Length 547;

Best Local Similarity 25.9%; Pred. No. 4.7e-08;

Matches 67; Conservative 43; Mismatches 85; Indels 64; Gaps 12;

QY 92 LSTDEVEKLEDFQOEFSPSRKGLV-LQLIQSYQRMPCGNAMVGRVAYKRHLVTMD 150

Db 282 INVDEEAQSPVAEQAEULP--EGLIKLQL-----ATYK---LIVDK 319

QY 151 LGTL-----YQGNLNDQVMNMYGDLVMDTVPEKVHFFNSFFYDKLRT-----KGYD---- 197

Db 320 TCSLQEDICYPTFRFLOQIIS-----SNOISADCHFFNTFYFKKLSDAVYKGNKDQAF 373

QY 198 --GVKRWTKNVDIFNKELLIPILHLEVHSLISV-----DVRRTITYFDSQRTLNRR- 248

Db 374 FVFRFRWWKGLDFKAYIPIPIHEDLHWSLVVICIPDKRDESLTILHDSLGLHSRKS 433

QY 249 CPKIAKYLOAEAVKKDRDLDF-----HGWKGYEK-----MNVARQNNDSDCCGAFVL 295

Db 434 IVENVRKFLKDEWNYLNQDDYSLDLPISKEVKNLPRRISEAVVQVPQKNDFDCGPFVL 493

QY 296 QYCKHLALSQFFSFTQODM 314

Db 494 FFIKRIEEAPQRLKRKDL 512

RESULT 10

E71424

hypothetical protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

A:Variety: Columbia

C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 11-Jan-2002

A:Accession: E71424

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; G  
avanagh, T.; Hempel, S.; Kottler, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;  
C.; Chalwatzis, N.

A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t

A:Reference number: A71400; MUID:98121113; PMID:9461215

A:Accession: E71424

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-398 <BEV>

A:Cross-references: GB:Z97339; NID:g2244901; PID:g2244945

C:Genetics:

A:Map position: 4COP9-4G3845

Query Match 10.7%; Score 190; DB 2; Length 398;

Best Local Similarity 24.0%; Pred. No. 5.5e-08;

Matches 69; Conservative 50; Mismatches 104; Indels 64; Gaps 10;

QY 9 GTIWARGAHLAPPDASILISNVCISGD-----HVAQELPQSGDLGMAEAEERPGEK 60

Db 125 GVTDVKKGFNFRYEDTSMMLDSLDRDNDASSLEAYRKLMQS-----AEKRNSKLEA 179

QY 61 AGQHSPLREEHVTVCQSILDEFQTYGSLIPSTDEWKEKLEDFQOEFSTPSRKGLVLQ 120

Db 180 LGFEIVLNEKKL-----SLLQSRPKTVKRVVPREPF-----IP 215

QY 121 LIQSYQRMPCGNAMVGRVAYKRHLV-----TMDLGLTYGONWLNDOVMNMYGD 170

Db 216 LTEDDEA-----EYVRAFSGNRNRKVLATHENSINIDITGEVLQCLTPSAWLNDVINVYLE 271

QY 171 LVMDTVPE-----KVHFFNSFFYDKL-RTKGYD--GVKRWTK-----NVDIFNKELLIP 217

Db 272 LKERETREFPKYKLYCHYNTFFYKLVSDSGYNFKAVRWFTQRLGLYALIDCDMIFVP 331

QY 218 IHLEVHWSLISVDVRRRTITYFDSQRTLNRCRKPHIAKYLOAEAVKK 264

Db 332 IHGCVHTLAVINNRESKLLYLDLSNGVDPMILNALAKYMGDEANEK 378

RESULT 11

T04979

hypothetical protein T16L1.110 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999

C:Accession: T04979

R:Bevan, M.; Obermaier, B.; Deutschenbaur, S.; Piravandi, E.; Hoheisel, J.; Mewes, H.

submitted to the Protein Sequence Database, November 1998

A:Reference number: Z15393

A:Accession: T04979

A:Molecule type: DNA

A:Residues: 1-710 <BEV>

A:Cross-references: EMBL:AL031394

A:Experimental source: cultivar Columbia; BAC clone T16L1

C:Genetics:

A:Map position: 4

A:Introns: 21/3; 97/3; 123/1; 141/1; 212/2; 279/2; 298/1; 327/2; 407/2; 518/2; 626/2

A:Note: T16L1.110

Query Match 9.2%; Score 163; DB 2; Length 710;

Best Local Similarity 31.7%; Pred. No. 2.1e-05;

Matches 32; Conservative 21; Mismatches 32; Indels 16; Gaps 2;

QY 145 VLTMDLGLTYGONWLNDOVMNMYGDLVMDTVPEK-----VHFNSFFYDKLRT----- 193

Db 187 VVRKQDIELLKPRRFINDTIIDFYIKYLNRIKSPKRGSRGFHFNCFFRKLANKDGTGS 246

QY 194 -----KGYDGVKRWTKNVDIFNKELLIPILHLEVHWSLISV 229

Db 247 TCGGREAYQRVKQTKNVDLFKDYIFIPINCSFHWSLVII 287

## RESULT 12

T48800  
SMT4 related protein [imported] - Neurospora crassa  
N:Alternate names: protein 15E6.80  
C:Species: Neurospora crassa  
C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 05-May-2000  
C:Accession: T48800  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24541  
A:Accession: T48800  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1240 <SCH>  
A:Cross-references: EMBL:AL353822; GSPDB:GN00112; NCSP:15E6.80  
A:Experimental source: cosmid contig 15E6; strain 74  
C:Genetics:  
A:Gene: NCSP:15E6.80  
A:Map position: 2  
A:Introns: 8/3; 358/2

Query Match 8.8%; Score 156; DB 2; Length 1240;

Best Local Similarity 18.2%; Pred. No. 0.00016;

Matches 86; Conservative 50; Mismatches 136; Indels 200; Gaps 14;

QY 27 LLSNVCSIGDHVAQELFOGSDGLGMAEAEERPCEKAGOHSPLEEHVTCVQSILDEFLQTY 86

Db 349 LSNVEMNSATLQKTYE-----NLMTLKK--ATSGSPSPLANSRVRTTSALD----- 396

QY 87 GSLIPLSTDEVVEKLEIDFQOEFSTPS-----RKGLVLQLIQSYQRMPCGNAMVRGFRV-- 139

Db 397 ---VEGSRNMAFEPAGLIAQAATAGSPTASTRRPRVLDVTLSSQQALSNOYEHRSFEVEA 453

QY 140 -----YKRVHVTMDDLGLTLY 155

Db 454 PYORSRRNRDPTVMVDAPPISSAPVSRWTEHSEWKNRMPLVYHTSVDKDDIPRLD 513

QY 156 GQNLNDQVMNMYGDLVM-----DTVPEKRVHFNFSFYDKLRTK-----GYDGVKRW 202

Db 514 EQCLNDNLG--FGLRYLFEFPGRHDELKRVYVHNTFFYELKPAKSKOINYGKVGW 572

QY 203 TKNVDIFNKELLIPHLVHVH----- 224

Db 573 TSKVLLSYDIIVPVNVEYHWWVAIICNPKGLDPNHRSTNSSTSGTETSNSSTESK 632

QY 225 -----SLISV----- 229

Db 633 SNGNIEKSDVEMIDICEQARDQGEIKAQACGSDVRIETDSNEADSAGQDVVDLVAD 692

QY 230 ----DVRRR-----TITYFDSORTLNRRCPKHIAKYLOAEAVKKDRLDF 269

Db 693 DTVDHLRELKGIITKLPPKDPSEETKILTDSMGNSHYPAVALKKYLMAEFEDKKQTKI 752

QY 270 HOGWK--GYFKMVARQNNDSCGAFVLYQYCKHLALSOPFSFT-----QODMP 315

Db 753 KDLPKQIGIKATNIPQNNFSDCGVYLLGYIOEF-VKQPDRAHCLMQREKP 803

## RESULT 13

T26650  
hypotheetical protein Y38A8.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T26650  
R:Hallsworth, K.  
A:Description: The sequence of C. elegans cosmid Y38A8.  
A:Reference number: Z20251  
A:Accession: T26650  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-920 <HAL>

A:Cross-references: EMBL:U55855; PIDN:AAA98019.1; GSPDB:GN00020; CESP:Y38A8.3  
A:Experimental source: strain Bristol N2; clone Y38A8

C:Genetics:

A:Gene: CESP:Y38A8.3

A:Map position: 2

A:Introns: 30/3; 58/3; 91/3; 182/3; 258/1; 288/2; 296/3; 339/3; 437/1; 505/3; 527/3;

Query Match 7.7%; Score 136; DB 2; Length 920;

Best Local Similarity 22.8%; Pred. No. 0.005;

Matches 65; Conservative 36; Mismatches 88; Indels 96; Gaps 13;

QY 87 GSLIPLS-----TDEVVEKLEIDFQOEFSTPSRKGLVLQLIQSYQRMPCGNAMVRGFRV 139

Db 523 GSCLPSTPTAALPPTRPVVEKIPP-DTQLFTFP-----PSGSCCTGMDP 565

QY 140 AYKRHVLTMDDLGLTLYGQNLNDQVM---NMVGDVMDTVPEKVVHFNFSF----- 187

Db 566 V-----VLLVKDKITLDRKEFLDSDVNAFMLNYIAFMSLSMKSVHMCMTFLFVNTRL 621

QY 188 ----YDKLRTKGYDGVK-----RWTKNVDIFNKELLIPHLVHWSLISV----- 229

Db 622 PPLCFSKRPIEPEHIKIVKDNCPVLRWTRKFDVLAKDYIIPINEDLHVLVIIVNS 681

QY 230 -----DVRRT-----ITYFDSORTL-----NRRCPKHIKYLQAEAVKKDRLDFHQ 271

Db 682 GAIVDMSNEEASRAAPKCVIVFFDPLSGLDPSKKNHMC--HCIKIYLAQLYENTKA---P 736

QY 272 GWKGVFK-----MVARQNNDSDCGAFVLYQYCKHL 301

Db 737 GWKGVFK-----MVARQNNDSDCGAFVLYQYCKHL 301

RESULT 14

T19877

hypotheetical protein C41C4.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T19877

R:Burton, J.

submitted to the EMBL Data Library, February 1995

A:Reference number: Z19190

A:Accession: T19877

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-342 <WIL>

A:Cross-references: EMBL:Z48045; PIDN:CAA98104.1; GSPDB:GN00020; CESP:C41C4.6

A:Experimental source: clone C41C4

C:Genetics:

A:Gene: CESP:C41C4.6

A:Map position: 2

A:Introns: 50/1; 110/3; 144/3; 327/3

Query Match 7.0%; Score 124.5; DB 2; Length 342;

Best Local Similarity 19.6%; Pred. No. 0.012;

Matches 63; Conservative 64; Mismatches 138; Indels 57; Gaps 13;

QY 23 DASILISVCSIGDHVAQELFOGS---DLGMAEAEERPCEKAGOHSPLEEHVTCVQSIL 79

Db 23 DVSMDREDDLFRMGSYNSQGYADGTHLDGSGIEEDTSSGSDNDQHMDFEEDDFDMESSMT 82

QY 80 DEFLOTGYLSTLSTDEVVEKLEIDFQOEFSTPSRKGLVLQLIQSYQRMPCGNAMVRGFRV 139

Db 83 ED-----LVDEDEDEEEDENDEWTKRTDNQNS-----VAYYAAM---EMLR-IRF 126

QY 140 AYKRHVLTMDDLGLTLYGQNLNDQVMNMY-GDLVMDTVPE-----KVHFTNSFYD---KLR 192

Db 127 PFQSTAIRISDFCCLEQKOLLNDTIDFVNLHIVEHVLVPSDNGSNVTPLPFIHNLRL 186

QY 193 TKGYDG-----VKRWTKNVDFNKELLIPHLVHWSLISV-----D 230

Db 187 QHAFDSEDEKMSDEQKMDLKFGLDHFVADFQDFYIVVPVNEWEHSLAVICHPTT 246

QY 231 VRRRTITYFDSQRTLNRRCPKHIKYLQAEAVKKDRLDFHQGWKGVFKM-----NVARQ 284

```

      : ||: |||| | : :|| : | : | : | : | : | : |
247 AQARTV-IFDSQLTADLNNLQNMTLII--ESPMKSYEKRTGNAMFFPLCILPQRMFQ 303
      : ||: |||| | : :|| : | : | : | : | : | : |
285 NNDSDCGAFVLOYCKHLALSQP 306
      : |||| : : : ||
304 TNNFDGGIFIAEFARFLLSPP 325

```

RESULT 15

T32915  
hypothetical protein K02F2.4 - *Caenorhabditis elegans*.  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T32915  
R:Maggi, L.; Goela, D.  
submitted to the EMBL Data Library, January 1998  
A:Description: The sequence of *C. elegans* cosmid K02F2.  
A:Reference number: Z21246  
A:Accession: T32915  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-311 <MAG>  
A:A:Cross-references: EMBL:AF043699; FIDN:AA897567.1; GSPDB:GN00019; CESP:K02F2.4  
A:Experimental source: strain Bristol N2; clone K02F2  
C:Genetics:  
A:Gene: CESP:K02F2.4  
A:Map position: 1  
A:Introns: 16/3; 72/2; 103/3; 139/2; 203/2; 238/2; 273/3  
C:Superfamily: *Caenorhabditis elegans* hypothetical protein K02F2.4

Query Match	6.7%;	Score 118.5;	DB 2;	Length 311;
Best Local Similarity	19.8%;	Pred. No. 0.034;		
Matches 51;	Conservative 35;	Mismatches 80;	Indels 91;	Gaps 10;
QY	142	KRHLVTMDLLTGLYGNW	-----LNDQVANNMYG-----	DLVMDTVPKVFHFNSF 186
		:	:	:
Db	28	RRHIIP-----PLFHNGWVG	NRDRTLLNDTIEFYMCDMR	LEVDFEDTRASSHVHFSF 82
QY	187	EYDKLAT-----	KGVDGKRWTKNVDIP-NKELL	LPIHLE-VVHSLI 227
		:	:	:
Db	83	FLPKIKTCFKDENPRRSELANH	YNNRFRPSKNDAAETFLUKD	ILLPIVHLDKPKHFLV 142
QY	228	SV-----DVRRTITY	FDQSORTLNNR-----	CPK----- 251
		:	:	:
Db	143	IVHNPAGVRRISDVNILLATN	ATNVKSRRLSRRTIGHVNC	DENAGECRIIMDSILVHSKY 202
QY	252	-----HIAKYLOAEAVK	KDRLDFHOGWGYFKMNVAR	QONNDSDCGAFVLOYC 298
		:	:	:
Db	203	REVIDKTHDSTFHIRILWLL	MSAAATDVDMCTFRFRVVC	QKLPQQRKNSVDGIFMMAFA 262
QY	299	KHLALSQPFSTQODMP	315	
		:	:	
Db	263	EYFT---KYNTAWOSLP	276	

Search completed: December 16, 2002, 20:15:46  
Job time : 259 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 19:28:45 ; Search time 207 Seconds  
(without alignments)  
66.522 Million cell updates/sec

Title: US-09-848-852A-3  
Perfect score: 1772  
Sequence: 1 MYSQRFWGTIWARRAHGA.....DMPKLRQIYKELCHCKLTV 332

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	319.5	18.0	568	ULP1_SCHPO	Q42957 caenorhabdi
2	268.5	15.2	659	YRW3_CAEEL	Q09353 schizosacch
3	222.5	12.6	621	ULP1_YEAST	Q02724 saccharomyc
4	205.5	11.6	652	ULP2_SCHPO	Q13769 schizosacch
5	196	11.1	1034	ULP2_YEAST	P40537 saccharomyc
6	170.5	9.6	984	SUS2_HUMAN	Q9bqf6 homo sapien
7	143	8.1	1112	SUS1_HUMAN	Q9gzrl homo sapien
8	124.5	7.0	342	YQ66_CAEEL	Q09275 caenorhabdi
9	105	5.9	1272	Y428_METUA	Q60287 methanococc
10	97	5.5	790	V90K_AWVLE	P03593 alfalfa mos
11	92.5	5.2	273	VPRP_ASEB7	Q00946 african swi
12	90	5.1	727	YR22_CAEEL	Q09639 caenorhabdi
13	89	5.0	608	HTPG_CAMEE	Q9phz3 campylobact
14	87.5	4.9	273	VPRP_ASEF2	Q65228 african swi
15	87.5	4.9	694	DNJ1_MYCLE	O33102 mycobacteri
16	87.5	4.9	2034	FER1_CAEEL	Q17388 caenorhabdi
17	87.5	4.9	4344	DYHC_EMENI	P45444 emericeella
18	86.5	4.9	382	HUGA_APIME	Q08169 apis mellif
19	86.5	4.9	484	R554_BRAJA	P30332 bradyrhizob
20	86	4.8	824	TG37_HUMAN	Q13099 homo sapien
21	85.5	4.8	538	Z155_HUMAN	Q12901 homo sapien
22	85	4.8	300	HTPG_MOUSE	Q9wtm0 m geranylge
23	85	4.8	607	HTPG_FUSNN	Q8rgh4 fusobacteri
24	85	4.8	735	RN5A_MOUSE	Q05921 mus musculu
25	85	4.8	802	PAC_ARTV1	P31956 arthrobacte
26	85	4.8	864	RA50_SULSO	Q97wh0 sulfolobus
27	84.5	4.8	562	SIR2_YEAST	P06700 saccharomyc
28	84.5	4.8	626	HTPG_BACSU	P46208 bacillus su
29	84.5	4.8	653	YJ04_YEAST	P39531 saccharomyc
30	84	4.7	402	ELAD_ECOLI	Q47013 escherichia
31	84	4.7	476	MPBP_NEUCR	P11913 neurospora
32	84	4.7	731	NGP1_HUMAN	Q13823 homo sapien
33	83.5	4.7	266	Y209_METUA	Q57662 methanococc

34	83.5	4.7	294	1	GGPP_BOVIN	P56966 b geranylge
35	83.5	4.7	4725	1	DYHC_DICDI	P34036 dictyosteli
36	83	4.7	366	1	MT52_METJA	Q58600 methanococc
37	83	4.7	503	1	FUTA_DROME	Q9vul9 drosophila
38	83	4.7	2386	1	RAD3_SCHPO	Q02099 schizosacch
39	82.5	4.7	420	1	YFJD_HAEIN	Q57017 haemophilus
40	82.5	4.7	459	1	YLD8_STRCO	O86835 streptomyce
41	82.5	4.7	625	1	RSC4_YEAST	Q02206 saccharomyc
42	82.5	4.7	1587	1	TOP2_PENCH	Q9y8g8 penicillium
43	82	4.6	353	1	LICH_PSEGL	Q05490 pseudomonas
44	82	4.6	803	1	SWI6_YEAST	P09359 saccharomyc
45	81.5	4.6	300	1	GGPP_HUMAN	O95749 h geranylge

ALIGNMENTS

RESULT 1  
ULP1\_SCHPO  
ID ULP1\_SCHPO STANDARD; PRT; 568 AA.  
AC Q42957;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ubiquitin-like-specific protease 1 (EC 3.4.22.-).  
GN ULP1 OR SPBC1967.09.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesti D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe."  
Nature 415:871-880(2002).  
[2]  
FUNCTION, AND GENE NAME.  
Watts F.;  
Submitted (MAR-2001) to the SWISS-PROT data bank.  
CC -!- FUNCTION: A CYSTEINE PROTEASE THAT PROTEOLYTICALLY REMOVES THE  
CC C-TERMINUS OF PMT3.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C48.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; A021839; CAAL17063.1; -  
HSP; Q02724; LEUV.  
MEROPS; C48.002; -  
InterPro; IPR003653; SUMO\_protease.  
Pfam; PF02902; Peptidase\_C48; 1.  
PROSITE; PS06000; ULP\_PROTEASE; 1.  
KW Hydrolyase; thiol protease.  
FT ACT\_SITE 463 BY SIMILARITY.  
FT ACT\_SITE 480 BY SIMILARITY.  
FT ACT\_SITE 527 BY SIMILARITY.  
SQ SEQUENCE 568 AA; 64940 MW; AAF9771440302D9F CRC64;

Query Match 18.0%; Score 319.5; DB 1; Length 568;  
Best Local Similarity 28.7%; Pred. No. 2.2e-18;  
Matches 90; Conservative 51; Mismatches 114; Indels 59; Gaps 9;  
QY 59 EKAGHSPLEEHVTCVQS-----ILDEPLQYGSILPLSTDEVVEKLEDFQQ----- 107  
Db 264 KSPHDSRKRASFVQSDQPPARNIVTSDIQEKSLLLRD-LKEQTESFQDNVEV 322  
QY 108 -----EFSTPSRKLGLVLIQSYQRMGPNAMVRGFRVAYKRH 144  
Db 323 DFLQLGLEISPPTRPKFIPPELPPDNARKALKYLQNSVSSSEPIITKFNIP----- 378  
QY 145 VLTWDDLTGLYQWLNDOVMYGDIV-----MDTVPEKVHFNFSFYKLRKTKGVDGV 199  
Db 379 -ITLKDHLTNRLQWLNDEINFYNNLISERSKIDSLPRVHGENTFTYSLQRGGYAGV 437  
QY 200 KRWTK--NVDIFNFKELLIPILHLEVHWSLISVDVRRRTITVFDQSORTLNRCPKHIKYL 257  
Db 438 RRAKARVNIADMDAIFVHLVDVHMCMAVINKRFEWDSLAGSPGVFLLRDYY 497  
QY 258 QAEAVKDRDLDFHOGWGYFKMVARQNDSDCGAFVLQYCKHALSQPFFSQDQMPKL 317  
Db 498 IAET--KGAVDV--SDWENFMDNDSRQNGHDCGVACKTAECVSRNVPVQFSQNDMPEL 554  
QY 318 R-----RQIY 322  
Db 555 RIKMAASIIDAQIY 568

RESULT 2  
YRW3\_CAEEL  
ID YRW3\_CAEEL STANDARD; PRT; 659 AA.  
AC Q09353;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein T10F2.3 in chromosome III.  
GN T10F2.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RC SEQUENCE FROM N.A.  
RA STRAIN=Bristol N2;  
RA Miller N.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RA Waterston R.;  
RC Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
RL -1- SIMILARITY: TO YEAST LPB11C AND S.POMBE SPBC19G7.09.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U23412; AAK21468.1; -  
HSP; Q02724; LEUV.  
MEROPS; C48.002; -  
WormPep; T10F2.3; CE26951.  
InterPro; IPR003653; SUMO\_protease.  
Pfam; PF02902; Peptidase\_C48; 1.  
DR PROSITE; PS06000; ULP\_PROTEASE; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 659 AA; 75304 MW; 4F66F0082D470EB7 CRC64;

Query Match 15.2%; Score 268.5; DB 1; Length 659;  
Best Local Similarity 26.4%; Pred. No. 3.6e-14;  
Matches 75; Conservative 53; Mismatches 79; Indels 77; Gaps 10;  
QY 27 LISNVCISG---DHVAQELFQSGDLGMAEAEPRGKAGQHSPLREHVTVCVQSILDFEL 83  
Db 393 ILSQINSLGIGSAYRGPRYQNS-----YQLSKQKEDKLLEEARIREGHS----- 438  
QY 84 QTVGSLPLSTDEVVEKLEDFQOEFTPSRKLGLVLIQSYQRM----- 128  
Db 439 QTRG-----DRLEDV-----RKRLQGLAIAPKPKVKKKVDDEMALPDAA 478  
QY 129 -----PCNAMVRGFRVAYKRHLVTMDLGLTYGNLNDQVMNMYGDLVM--- 173  
Db 479 DALVERAMSGNNEQVDAFSIQ-----ICKKDLATLSGLHNLNDEIINFYLIQLICDRS 533  
QY 174 --DTPVEKVFHFNFSFYDKLRTKGYCGVKKRWTKNVDIFNFKELLIPILHLEVHWSLISVDV 231  
Db 534 NGDSKYPKIYAFNTFFYSNIVSGYASVKKRWTKVDIFAEDIVLVPVHLGHHMCMAVIDM 593  
QY 232 RRTTIYFDSQRTLNRCPKHIKYLQAEV--KKRLDFHQGW 273  
Db 594 GKKEIFYDSDYDGTAVLPALRGYLEAEASLDKKTKAMNF-SGW 636

RESULT 3  
ULP1\_YEAST  
ID ULP1\_YEAST STANDARD; PRT; 621 AA.  
AC Q02724;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ubiquitin-like-specific protease 1 (EC 3.4.22.-).  
GN ULP1 OR YPL020C OR LPB11C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=97313271; PubMed=9169875;  
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,  
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,  
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,  
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,  
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,  
RA Kompe C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,  
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,  
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,  
RA Scherens B., Schramm S., Schroeder M., Sidu A.M., Tettelin H.,  
RA Urrestazu L.A., Ushinsky S., Vierdeels F., Vissers S., Voss H.,  
RA Walsh S.V., Wambutt R., Wang Y., Wedler H., Wedler H., Winnert E.,  
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";

ULP2_SCHPO	STANDARD;	PRT;	552 AA.
ID	ULP2_SCHPO		
AC	013769;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Ubiquitin-like-specific protease 2 (EC 3.4.22.-)		
DN	ULP2 OR SPAC17A5.07C.		
OS	Schizosaccharomyces pombe (Fission yeast).		
OS	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;		
OC	Schizosaccharomycetes.		
OX	NCBI_TaxID=4896;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=972;		
RC	MEDLINE=21848401; PubMed=11859360;		
RA	Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,		
RA	Sgueros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,		
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,		
RA	Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,		
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,		
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,		
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,		
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,		
RA	Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,		
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,		
RA	Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,		
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,		
RA	Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,		
RA	Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,		
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,		
RA	Borzyn K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,		
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,		
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Laureau V., Mottier S.,		
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,		
RA	Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,		
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,		
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,		
RA	Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,		
RA	Shpakovski G.V., Ussery D., Barrall B.G., Nurse P.,		
RL	"The genome sequence of Schizosaccharomyces pombe.";		
RL	Nature 415:871-880(2002).		

```

[2]
nature 413:871-880(2002).
RN      GENE NAME.
RP      Watts F.;
RA      Submitted (MAR-2001) to the SWISS-PROT data bank.
RL      -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C48.
RC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z98849; CAB11507.1; -.
CC MEROPS; C48. UPW: -.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02502; Peptidase_C48; 1.
DR PROSITE; PS00600; ULP_PROTEASE; 1.
DR Hydrolase; Thiol protease.
DR ACT_SITE 454 454
DR ACT_SITE 558 558 BY SIMILARITY.
FT ACT_SITE 558 558 BY SIMILARITY.
FT SEQUENCE 652 AA; 73760 MW; 5A4F5D5DE97BD05E CRC64;
SQ
Query Match 11.6%; Score 205.5; DB 1; Length 652;
Best Local Similarity 26.0%; Pred. No. 4.4e-09;
Matches 60; Conservative 36; Mismatches 72; Indels 63; Gaps 63;

```

## RESULT 4





```
FT DOMAIN 694 984
FT ACT_SITE 794 794
FT ACT_SITE 926 926
FT ACT_SITE 926 926
FT CONFLICT 275 275
FT CONFLICT 330 330
FT CONFLICT 330 330
FT CONFLICT 546 546
FT CONFLICT 747 761
FT CONFLICT 828 828
FT CONFLICT 881 881
FT CONFLICT 881 881
SQ SEQUENCE 984 AA: 112264 MW: 317E8B7B4128EDID CRC64;

Query Match 9.6%; Score 170.5; DB 1; Length 984;
Best Local Similarity 22.6%; Pred. No. 5.1e-06;
Matches 64; Conservative 41; Mismatches 67; Indels 111; Gaps 11;

QY 146 LTMDDLTGLYGNWLNQVMNMYGDLV-----DTPVKVHFFNFFFDKLRTK----- 194
DB 696 VTNEDELECEGEFLNDVIIDFYLYLILEKASDELVERSHIFFSFFYKCLTRKENNLTE 755
QY 195 -----GYDGVKRWTKNVDIENKELLIPILHLEVHWSL----- 226
DB 756 DNPNLMSAQRHKKVTRTHININENKDYIFVPVNESSHLYLAVICFPWLEAVYEDFPQ 815
QY 227 -----ISVDVRRR---TITVDSQRT-LNRRCPKHIAK----- 255
DB 816 TVSQSQQAQSQNDKNTIDNLTSTLSAEDSOSTESNMSPKMKCRPCILILDSL 875
QY 256 -----YLQAEVAKDRIDFHQGWKGFKMN-----VARQNDSDCGAFV 294
DB 876 KAASQNTVQNLRLEYEWEVK--LKTH---RQFSKTNMVDLCPKPKQDNSSDCGVYL 930
QY 295 LOYCKHLALSOPFTQODMP-----KLRRQIYKEL 325
DB 931 LOYVESFKDPIVNF---ELPIHLEKWPFRHVIKTKREDIREL 970

RESULT 7
SUSL_HUMAN
ID SUSL_HUMAN STANDARD; PRT; 1112 AA.
AC Q9GZRL: Q9UJV5; O94891;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SUMO-1-specific protease 1 (EC 3.4.22.-) (Sentrin-specific protease
DE SENP6) (Protease FKSG6).
GN SUSP1 OR SSP1 OR KIAA0797.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=20261527; PubMed=10799485;
RA Kim K.I., Baek S.H., Jeon Y.-J., Nishimori S., Suzuki T., Uchida S.,
RA Shimbara N., Saitoh H., Tanaka K., Chung C.H.;
RT "A new SUMO-1-specific protease, SUSP1, that is highly expressed in
RT reproductive organs."
RL J. Biol. Chem. 275:14102-14106(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20267842; PubMed=10806345;
RA Yeh E.T., Gong L., Kamitani T.;
RT "Ubiquitin-like proteins: new wines in new bottles."
RL Gene 248:1-14(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Wang Y.-G.;
RT "Identification of FKSG6, a novel protein with protease activity."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
```

```
RP SEQUENCE OF 29-1112 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 5:277-286(1998).
CC -!- FUNCTION: Protease that release SUMO-1 from its precursor
CC sequence.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Highly expressed in reproductive organs, such
CC as testis, ovary, and prostate.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF196304; AAF04852.1; -
CC EMBL; AF307849; AAG29831.1; -
CC EMBL; AF306508; AAG30253.1; -
CC EMBL; AB018340; BAA34517.1; -
CC MEROPS: C48.004; -
CC MIM: 605003; -
CC InterPro: IPR003653; SUMO-protease.
CC Pfam: PF02902; Peptidase_C48; 1.
CC PROSITE: PS06000; ULP_PROTEASE; 1.
KW Hydrolase; Thiol protease; Ubl conjugation pathway.
FT DOMAIN 666 1112
FT ACT_SITE 765 765
FT ACT_SITE 917 917
FT ACT_SITE 1030 1030
FT CONFLICT 121 121
FT CONFLICT 1043 1043
FT CONFLICT 1106 1106
SQ SEQUENCE 1112 AA; 126144 MW; B6E53875C3833A46 CRC64;

Query Match 8.1%; Score 143; DB 1; Length 1112;
Best Local Similarity 30.8%; Pred. No. 0.001;
Matches 32; Conservative 23; Mismatches 29; Indels 20; Gaps 3;

QY 146 LTMDDLTGLYGNWLNQVMNMY-GDLVMDTV-----PEKVHFFNFFFDKLR----- 191
DB 668 VTNEDELECEGEFLNDVIIDFYLYLILEKASDELVERSHIFFSFFYKCLTRKENNLTE 727
QY 192 -----RTKGYDGVKRWTKNVDIENKELLIPILHLEVHWSLISV 229
DB 728 TTNLSIQQRHGRVKTWTRHVDIFEKDFIFVPLNEAAHFWLAV 771

RESULT 8
YQG6_CAEEL
ID YQG6_CAEEL STANDARD; PRT; 342 AA.
AC Q09275;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 35.7 kDa protein C41C4.6 in chromosome II.
GN C41C4.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Burton J.;
```

```
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z48045; CAA88104.1; -.
CC DR MEROPS; C48.UPW; -.
CC DR WormPep; C41C4.6; CE01521.
CC DR InterPro; IPR003653; SUMO_protease.
CC DR Pfam; PF02902; Peptidase_C48; 1.
CC DR PROSITE; PS06000; ULP_PROTEASE; 1.
CC DR Hypothetical protein; Hydrolase; Thiol protease.
CC KW SEQUENCE 342 AA; 39503 MW; 80ABF2AECCAECAD4 CRC64;
CC -----
Query Match 7.0%; Score 124.5; DB 1; Length 342;
Best Local Similarity 19.6%; Pred. No. 0.007;
Matches 63; Conservative 64; Mismatches 138; Indels 57; Gaps 13;
CC
QY 23 DASILISNVCISGDHVAQELFOGS---DLGMAEEAERPGEKAGOHSPLRREHVTVCQSIL 79
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
23 DVSMREDDLFMRGYSNSQGYADGTHLDGSGIGEEDETSSGNSQNDQHMDFEEDFDMESSMT 82
QY 80 DEFLOTVGSILPSTDEWEKLEIDFOQEFSTPSRKLVLQLIQSQVQRMGNAMVRGRFV 139
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
83 ED-----LVDEDEEEDDEEDNWTNKRDTNQNS-----VAYAAAM---EMLR-IRF 126
QY 140 AYKRHLVTMDLGLTYGQNLNDQVMNY-GDLVMDTPVE---KVHFFNFFYD--KLR 192
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
127 PFQSIATRISDFCCLQKQLDNLNTMIDFNLHIVEHVLPDSNGSNVTVLPSIFVHNLNLSR 186
QY 193 TKGYDG-----VKRWTKNVDIFNKELLIPILHLEVHWSLISV-----D 230
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
187 QHAPDSEDEKMSDEQKMDLKFGLHDFVADFDLQDFDIYVVPVNEWEHWSLAVICHPT 246
QY 231 VRRITVYFDSQRLNRCRPHIAKYLAQAEAVKKDRDLDFHQGWKGYFKM-----NVARQ 284
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
247 AQARTV-IFDSQTLADLNNLNQNTATLI--ESFMKYSYKRTGNAMPFPLPILQPMRPPQ 303
QY 285 NNDSDCGAFVLQCKKHLASQP 306
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
304 TNNFDCGIFTAEARRELLSFP 325
CC -----
RESULT 9
Y228_METJA STANDARD; PRT; 1272 AA.
ID Y228_METJA
AC Q60287;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJEC128.
GN MJEC128
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Buit C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
CC -----
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii."
CC Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; L771118; AAC37099.1; -.
CC DR TIGR; MJEC128; -.
CC DR Hypothetical protein; Complete proteome.
CC KW SEQUENCE 1272 AA; 149438 MW; 88E7AE8C48629DEA CRC64;
CC -----
Query Match 5.9%; Score 105; DB 1; Length 1272;
Best Local Similarity 25.9%; Pred. No. 1.4;
Matches 42; Conservative 13; Mismatches 39; Indels 68; Gaps 6;
CC
QY 144 HVLTMDDLGLTYGQNLNDQVMNYGDLVMDTPVEKVFHFFNSFFYDKLRTKGYDGVKRW 203
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
570 HYLTV---IGTY-----DQVENLNDLCYDT-----YKGEK--- 596
QY 204 KNVDIENKELLIPILHLEVHWSLISVDVRRITIT----- 237
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
597 --IQVANEDELLAID-EAHLVLIQEFKRRAITGVKKCIDRAGGCVLLTATPELINLNN 653
QY 238 -----YFDSQRLNRCRPHIAKYLAQAEAVKKDRDLDFHQGW 273
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
654 YPVIEVFEKDEKLFERCSTHIAKNIIGFECEVILLMFRQGW 695
CC -----
RESULT 10
V90K_AMVLE STANDARD; PRT; 790 AA.
ID V90K_AMVLE
AC P03593;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 90 kDa protein.
OS Alfalfa mosaic virus (strain 425 / isolate Leiden).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
OC Alfamovirus.
OX NCBI_TaxID=12322;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83220723; PubMed=6304618;
RA Cornelissen B.J.C., Brederode F.T., Veeneman G.H., van Boom J.H.,
RA Bol J.F.;
RT "Complete nucleotide sequence of alfalfa mosaic virus RNA 2."
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X01572; CAA25728.1; -.
CC DR EMBL; AK02702; AAA46290.1; -.
CC DR PIR; A04201; WMFM9.
CC DR InterPro; IPR001788; RNA_dep_RNAPol2.
CC DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
CC KW SEQUENCE 790 AA; 89855 MW; B435C9EB9AA853C2 CRC64;
CC -----
Query Match 5.5%; Score 97; DB 1; Length 790;
Best Local Similarity 20.5%; Pred. No. 3.4;
Matches 52; Conservative 35; Mismatches 75; Indels 92; Gaps 10;
```



```
FT DOMAIN 191 455 PROTEIN KINASE.
FT DOMAIN 456 700 C-TERMINAL.
FT DOMAIN 54 175 RGS.
FT NP_BIND 197 205 ATP (BY SIMILARITY).
FT BINDING 220 220 ATP (BY SIMILARITY).
FT ACT_SITE 318 318 BY SIMILARITY.
SQ SEQUENCE 727 AA; 83361 MW; AA05308A9DA57D7 CRC64;

Query Match 5.1%; Score 90; DB 1; Length 727;
Best Local Similarity 19.7%; Pred. No. 11;
Matches 62; Conservative 50; Mismatches 122; Indels 80; Gaps 15;

QY 48 LGMAEAEPEKAGQSPHREHVTVCQSILDEFLQTYGSLPLSTDEVVEKLEDFQ 107
Db 15 MAMEKRSQPAARAKRVLDPD---PSVRSIMQKLEKSGDM-----KFDKIFNQ 61
QY 108 EF-----STPSKGLVLQIQSYQRM--PCGNAMVRGFRVAYKRHLVTMDL 152
Db 62 KLGFLLLKDYAENVSESPQIKFYEAKEYKMETPDERLTKA-REIYDHHIM-VEMLA 119
QY 153 TLGQGNLNDQVMNMGDLVMDTV-PEKVHVFNSFFYDKLRTKGYDGVKRWTKNVDIFNK 211
Db 120 --HANYSKESLQHVHLLKQNVPPDLFRHYVLEICDQLRG-----DIFQR 164
QY 212 ELLIPIHLEHVHWSLISVDVRRRTITYFDSQRTLN-----RCPK-HIAKYLQAEAVK 263
Db 165 FLESKDFTRECKWNLNN-QLTMNDFSVHRIIGRGEGEVYGCRAKDTGKMYAMKCLD 223
QY 264 KRLDFHQHKGVEK-----MNVARQNNDSDCGAFVLQYCK-----H 300
Db 224 KRKKMKQGGTALNEHIMLSLSTGDCPFIVCMTYAFQSPDKLC--FILDLMNGDGLH 281
QY 301 LALSQPFSTQDDM 314
Db 282 YHLSQHGVTQEOM 295

RESULT 13
HTPG_CAMJE STANDARD; PRT; 608 AA.
AC QSPH23;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chapterone protein htpg (Heat shock protein htpg) (high temperature protein G).
DE HTPG OR CJ0518.
GN Campylobacter jejuni.
OS Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karylishev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL139075; CAB75155.1; -.
DR HSSP: P02829; IAH8.
DR InterPro: IPR003594; ATPbind_ATPase.
DR InterPro: IPR001404; Hsp90.
DR Pfam: PF00183; HSP90; 1.
DR Pfam: PF02518; HATPase_C; 1.
DR PRINTS: PR00775; HEATSHOCK90.
DR SMART: SM00387; HATPase_C; 1.
DR PROSITE: PS00298; HSP90; 1.
KW Chapterone; ATP-binding; Heat shock; Complete proteome.
FT DOMAIN 1 332 A; SUBSTRATE-BINDING (BY SIMILARITY).
FT DOMAIN 333 536 B (BY SIMILARITY).
FT DOMAIN 537 608 C.
SQ SEQUENCE 608 AA; 69654 MW; 7FB40F4B365DF6B5 CRC64;

Query Match 5.0%; Score 89; DB 1; Length 608;
Best Local Similarity 21.1%; Pred. No. 11;
Matches 62; Conservative 45; Mismatches 95; Indels 92; Gaps 14;

QY 68 REBHVTVCQSIL--DEFLQTYGSLPLSTDEVVEKLEDFQ-----QEFSTPSRKG-- 116
Db 160 KEQGTSTITLYLKXDEFANAY-----KIESIIEKYSNHIQFIFMEKEEF-TPAKEGEE 212
QY 117 ---LVQLQTSQYQRMPCGNAMVRGFRVAYKRHLVTMDLGLTYGNLNDQVMNMGDLVM 173
Db 213 EGTEKIKISQINK---ANALWR-----MOKSLKAEDYERFYEQNFHDSNKKPLLY---L 260
QY 174 DTVPKRVHVFNSPF-----YDKLRTKGYDGVKRWTKNVDIFNKELLIPIHLEHVHWSL 226
Db 261 HTKSEKGLKLEYNLSLFFIPQNPAPDLFRVDYQSLGLKLVKRVFISDDKELLPTVLRVIRGI 320
QY 227 ISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQAEAVKVRDLDPHQGWKGYFKNNVAR--- 283
Db 321 IDVE-----DLPLN-----VSREILQENQILKGKIEASVKKILG 354
QY 284 -----QNNDSDCGAFVLQYCKHLALSQPF-----SFTQDDMPKLRRIQYK 323
Db 355 ELEKLNNDKE-----KYLSPFTKGVKLVKGLYGGGKDSLLKLMLYK 399

RESULT 14
VPRT_ASFM2 STANDARD; PRT; 273 AA.
AC Q65228;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Polypeptide processing peptidase (EC 3.4.22.-).
GN ORF16R.
OS African swine fever virus (isolate Malawi Lil 20/1) (ASFV).
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
OX NCBI_TaxID=10500;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94292916; PubMed=8021596;
RA Dixon L.K., Twigg S.R.F., Baylis S.A., Vydelingum S., Bristow C.,
RA Hammond J.M., Smith G.L.;
RT "Nucleotide sequence of a 55 kbp region from the right end of the
RT genome of a pathogenic African swine fever virus isolate (Malawi
RT LIL20/1)."
RL J. Gen. Virol. 7:1655-1684(1994).
CC -!- FUNCTION: Protease involved in the processing of the virus
CC polyproteins (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL; X71982; CAA50819.1; -;  
DR MEROPS; C48.050; -;  
KW Hydrolase; Thiol protease; Late protein.  
FT ACT\_SITE 168 187 BY SIMILARITY.  
FT ACT\_SITE 187 187 BY SIMILARITY.  
FT ACT\_SITE 232 232 BY SIMILARITY.  
SQ SEQUENCE 273 AA; 31493 MW; D6AD6DAC62C25F3B CRC64;

Query Match 4.9%; Score 87.5; DB 1; Length 273;  
Best Local Similarity 25.0%; Pred. No. 5.2;  
Matches 28; Conservative 21; Mismatches 44; Indels 19; Gaps 5;

QY 223 HNSLSIVVRR--TTTFDSQRTLNRRCPKHIKYLQAEVKKRDLDFHQGWKGFKNW 280

Db 168 HWAFVDMRGDCWSIEYFNS---AGNPPGPVIRWM--ERVQKQLKIHHTVTKTLAVTN 222

QY 281 VARQNDSDCGAFVLYQCK-----HLALSQPSFTQDMPKLRROIYK 323

Db 223 IRHRSQTECGPYLFYIRALDNVSYTHFISTR-----ITDENMYKPRTHLFR 271

#### RESULT 15

DNLJ MYCLE  
ID DNLJ\_MYCLE STANDARD; PRT; 694 AA.  
AC O33102;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA ligase (EC 6.5.1.2) (Polydeoxyribonucleotide synthase [NAD+]).  
GN LIGA OR Lig OR ML1705 OR MLCB637.10.  
OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier C., Churche C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;

RT "Massive gene decay in the leprosy bacillus.";

RL Nature 409:1007-1011(2001).

CC -!- FUNCTION: THIS PROTEIN CATALYZES THE FORMATION OF PHOSPHODIESTER  
CC LINKAGES BETWEEN 5'-PHOSPHORYL AND 3'-HYDROXYL GROUPS IN DOUBLE-  
CC STRANDED DNA USING NAD AS A COENZYME AND AS THE ENERGY SOURCE FOR  
CC THE REACTION. IT IS ESSENTIAL FOR DNA REPLICATION AND REPAIR OF  
CC DAMAGED DNA (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: NAD(+) + [deoxyribonucleotide](N) +  
CC [deoxyribonucleotide](M) = AMP + nicotinamide nucleotide +  
CC [deoxyribonucleotide](N+M).

CC -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT DNA LIGASE FAMILY.

CC -!- SIMILARITY: CONTAINS 1 BRCT DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL; Z99263; CAB16425.1; -;  
DR EMBL; AL583923; CAC30658.1; -;  
DR HSP; O87703; 1B04.  
DR Leproma; ML1705; -;  
DR InterPro; IPR001357; BRCT.  
DR InterPro; IPR004150; DNA\_ligase\_OB.  
DR InterPro; IPR001679; DNAligase.  
DR InterPro; IPR003583; HHH\_1.  
DR InterPro; IPR000445; Hhh.  
DR InterPro; IPR004149; Znf\_DNAligase\_C4.  
DR Pfam; PF00533; BRCT; 1.  
DR Pfam; PF00633; HHH; 1.  
DR Pfam; PF01653; DNA\_ligase\_N; 1.  
DR Pfam; PF03119; DNA\_ligase\_2BD; 1.  
DR Pfam; PF03120; DNA\_ligase\_OB; 1.  
DR ProDom; PD003944; DNAligase; 1.  
DR SMART; SM00292; BRCT; 1.  
DR SMART; SM00278; HHH1; 1.  
DR SMART; SM00532; LIGANC; 1.  
DR TIGRFAMS; TIGR00575; dnlj; 1.  
DR PROSITE; PS0172; BRCT; 1.  
DR PROSITE; PS01055; DNA\_LIGASE\_N1; 1.  
DR PROSITE; PS01056; DNA\_LIGASE\_N2; FALSE\_NEG.  
KW Ligase; DNA repair; DNA replication; NAD; Complete proteome.  
FT DOMAIN 613 677  
FT BINDING 123 123 AMP (BY SIMILARITY).  
SQ SEQUENCE 694 AA; 75912 MW; CFE712B6AFDAA9D5 CRC64;

Query Match 4.9%; Score 87.5; DB 1; Length 694;  
Best Local Similarity 23.1%; Pred. No. 17;  
Matches 67; Conservative 37; Mismatches 123; Indels 63; Gaps 13;

QY 45 GSDLGNA-----EERPRGKAGCOHSPREEHVTCVQSILDEFLQTYGSLPLSTDEVVE 99

Db 360 GSTVGQATLHNPSEVRKRGVLIGDTVVIRKA-----GDVPEVLGFPVD 403

QY 100 KLEDFPQOEF-----STPSRKGLVLQLIQSYQRMPCGNAMVGRFVAYKRHLVT 147

Db 404 -LRDGSEREFEVMTTCPEGCTTLAPEKEGDADIRCPNARSCPQLRERVFHVA-SRSALD 461

QY 148 MDGLTYLQCNWINDQVMNMGDLVMDTVPEKVFHFNFSFFYKLRTKGYDGVKRWTKNVD 207

Db 462 IQGLGYEAGVALIAAQVITSEGLF--TLTEKALLRTELFRNKAGELSANG-KRLLVNVD 518

QY 208 IFNKELLPIHLEVHWS-LISVDVR-----RRITVDSQRTLNRRCPKHIKY--- 256

Db 519 -----KAKTAPL-----WRVLVALSIRHVGPTAARALATEFGSDAILAASPQOLAAGEV 569

QY 257 ---LQAEAVKKRDLDFHQGWKGYFRKMNVARQNNDSDCGAFVLYQYKHLAL 303

Db 570 GTTIAAAVTEWFTVDWHRVIVNKRAGVYRMDERDTS--VLPTCEGLTI 617

Search completed: December 16, 2002, 19:58:11

Job time : 210 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2002, 20:15:50 ; Search time 299 seconds  
(without alignments)  
2500.547 Million cell updates/sec

Title: US-09-848-852A-3  
Perfect score: 1772  
Sequence: 1 MYSQREWTGWIARRGAHLA.....DMPKLRRIQYKELCHCKLTV 332

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-Q/cgn2.1/USPFO\_spool/US09848852/runat\_16122002\_132039\_19101/app-query.fasta\_1.519  
-DB=N.Geneseq\_101002 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09848852 -CGCN\_1.1.281 @runat\_16122002\_132039\_19101 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq\_101002.\*  
1: /SID32/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID32/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID32/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID32/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID32/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID32/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID32/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID32/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID32/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID32/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID32/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID32/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID32/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID32/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID32/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID32/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID32/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID32/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID32/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID32/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID32/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID32/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1751	98.8	1760	21	AAZ43782
2	1751	98.8	2080	22	AAI58425
3	1751	98.8	2206	22	AAF54831
4	1703	96.1	1507	22	AAO5541
5	1702	96.0	1934	22	AAO5502
6	1699	95.9	1804	22	AAI60211
7	1699	95.9	1804	22	AAI60212
8	1674	94.5	1933	24	ABQ54216
9	1642	92.7	2029	22	AAI58426
10	820	46.3	1578	22	AAH13708
11	784	44.2	870	22	AAI94288
12	755	42.6	1938	22	AAH99558
13	661	37.3	503	22	AAI84371
14	618	34.9	358	22	AAI84371
15	614	34.7	820	22	AAI93958
16	580	32.7	714	22	AAH03368
17	474	26.8	5318	22	AAF55961
18	456	25.7	2512	22	AAF54827
19	393	22.2	2570	23	ABV22135
20	393	22.2	2570	23	ABV27975
21	393	22.2	2795	22	AAF54830
22	393	22.2	3133	22	AAI93888
23	393	22.2	3133	22	AAH17779
24	353	19.9	650	21	AAC59569
25	312	17.6	1773	23	ABL05231
26	311	17.6	3963	23	ABL05230
27	310	17.5	9096	23	ABL17142
28	308	17.4	7494	23	ABL13550
29	307	17.4	595	20	AAV84693
30	307	17.3	4907	23	ABL13551
31	299	16.9	181	21	AAC08625
32	297	16.8	666	23	AAI71205
33	289	16.3	208	21	AAI45202
34	261	14.8	1859	21	AAC48444
35	259	14.6	1861	21	AAC40619
36	255	14.4	862	21	AAA02686
37	236	13.3	10715	23	ABL21052
38	236	13.3	10734	23	ABL21050
39	232	13.1	13525	23	ABL04986
40	215	12.2	695	22	AAO5540
41	205	11.6	631	20	AAV84692
42	196	11.1	33954	21	AAZ93712
43	190	10.7	501	22	AAK61344
44	173	9.8	3447	23	ABL16117
45	170	9.6	2618	24	ABL41323

ALIGNMENTS

RESULT 1

AAZ43782

ID AAZ43782 standard; cdna; 1760 BP.

XX AAZ43782;

AC AAZ43782;

XX 10-MAR-2000 (first entry)

DT Human fetal brain cdna clone vb7\_1.

DE Human; secreted protein; treatment; nutritional activity; cytokine;  
XX cell proliferation; cell differentiation; hematopoiesis regulation;  
KW tissue growth; activin; inhibin; chemotactic; chemokinetic; hemostatic;  
KW thrombolytic; anti-inflammatory; invasion suppressor; tumor inhibition;  
KW gene therapy; ds.

OS Homo sapiens.

XX

PN WO9955721-A1.  
 XX  
 PD 04-NOV-1999.  
 XX  
 XX 23-APR-1999; 99WO-US08504.  
 XX  
 PR 24-APR-1998; 98US-0082904.  
 PR 11-JUN-1998; 98US-0088994.  
 PR 12-JUN-1998; 98US-0089278.  
 PR 02-JUL-1998; 98US-0091647.  
 PR 24-AUG-1998; 98US-0097639.  
 PR 22-APR-1999; 99US-0097639.  
 XX  
 PA (ALPH-) ALPHAGENE INC.  
 XX  
 XX Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;  
 PI P-PSDB; AAY50911, AAY50912.  
 XX  
 DR WPI: 2000-052801/04.  
 DR P-PSDB; AAY50911, AAY50912.  
 XX  
 PT New polynucleotides encoding secreted human proteins, derived from  
 PT human fetal brain, adult skin, adult brain, adult heart, adult thymus  
 PT and adult aorta cDNA libraries.  
 XX  
 PS Claim 20a: Page 223; 282pp; English.  
 XX  
 CC This invention describes novel human secreted proteins which are encoded  
 CC by polynucleotides obtained from fetal brain, adult skin, adult brain,  
 CC adult heart, adult thymus and adult aorta cDNA libraries. The  
 CC polynucleotides and proteins are predicted to have biological activities  
 CC which would make them suitable for treating, preventing or ameliorating  
 CC medical conditions in humans and animals, although no supporting data  
 CC is given. Suggested activities include nutritional activity, cytokine  
 CC and cell proliferation/differentiation activity, immune stimulating  
 CC (e.g. as vaccines) or suppressing activity, hematopoiesis regulating  
 CC activity, tissue growth activity, activin/inhibin activity,  
 CC chemotactic/chemokinetic activity, hemostatic and thrombolytic activity,  
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumor  
 CC invasion suppressor activity, and tumor inhibition activity. The  
 CC polynucleotides are also stated to be useful for gene therapy.  
 CC AA243777-243808 represent the polynucleotides described in the invention  
 CC which encode the proteins represented in AAY50905-Y50947.  
 XX  
 SQ Sequence 1760 BP; 446 A; 444 C; 450 G; 420 T; 0 other;  
 XX

Alignment Scores:  
 Pred. No.: 1,18e-193 Length: 1760  
 Score: 1751.00 Matches: 331  
 Percent Similarity: 99.40% Conservative: 0  
 Best Local Similarity: 99.40% Mismatches: 1  
 Query Match: 98.81% Indels: 1  
 DB: 21 Gaps: 0

US-09-848-852A-3 (1-332) x AA243782 (1-1760)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20  
 DB 254 ATGACTCTGCCCAACGGTTTGGGGACAACTCTGGCCAGAGGGAGCGCAGCTTGGC 313  
 QY 20 aProToAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40  
 DB 314 ACCCCCTGATGCCAGCATCTCATACGAATGTGTGCAGCATCGGGAGCATGTGGCCCA 373  
 QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60  
 DB 374 GGAGCTTTTTCAGGGCTCAGATTGGCATGGCAGAGAGGAGGAGCGCTGGGGAGAA 433  
 QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValcInSerIleLeuAs 80  
 DB 434 AGCGGCCAGCAGACCCCTCGAGAGGAGCATGTGACCTGGGTACAGAGCATCTTGA 493  
 QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLy 100

Db 494 CGAATTCCTTCAAAACGTATGGCAGCCTCATACCCCTCAGCACTGATGAGTAGAGAA 553  
 QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGl 120  
 Db 554 GCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGCCCTGGTGTGCA 613  
 QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140  
 Db 614 GCTGATCCAGCTTACCAGGGATGCCAGCAATGCCATGCTGAGGGGCTCCGAGTGGC 673  
 QY 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160  
 Db 674 TTATAAGCGGCACGTGCTGACCATGGATGACTTGGGGACCTTGTATGACAGAACTGGCT 733  
 QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180  
 Db 734 CAATGACCAAGGTGATGAACATGTATGGAGACCTGGTGCATGGACACACATCCCTGAAAGGT 793  
 QY 180 lHisPhePheAsnSerPhePheTyrAspLysLeuAlaGThrLysGlyTyrAspGlyVally 200  
 Db 794 GCATTTCTTCAATAGTTTCTTCTATGATAAATCCGTCACCAAGGGTTATGATGGGTGAA 853  
 QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220  
 Db 854 AAGGTGGACCAAAACGTGGACATCTTCAATAAGGAGCTACTGTAATCCCATCCACCT 913  
 QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgArgThrIleThrTyrPheAs 240  
 Db 914 GGAGGTGCATTTGGTCCCTCATCTCTTGTATGTGGCGGACGCCACCATCACCTATTTTGA 973  
 QY 240 pSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260  
 Db 974 CTGCGACGCTACCTTAACCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAGA 1033  
 QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280  
 Db 1034 GCGCGTAAAGAAAGACCGCATGGATTTCCACAGGCTGGAAGGTACTTCAAAATGAA 1093  
 QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300  
 Db 1094 TGTGGCCAGCAGAAATAATGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1153  
 QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgGl 320  
 Db 1154 TCTGGCCCTGTCTCAGCCCATTCAGCTTCACCCAGCAGGACATGCCCAAACTTCGTCGGA 1213  
 QY 320 nIleTyrLysGluLeuGlyHisCysLysLeuThrVal 332  
 Db 1214 GATCTAAAGGAGCTGTGTCACTGCAAACTCACTGTG 1250

RESULT 2  
 AAI58425  
 ID AAI58425 standard; cDNA: 2080 BP.  
 XX  
 AC AAI58425;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 628.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX



PF 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI; 2001-442253/47.  
DR P-PSDB; AAM39269.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX Claim 1: SEQ ID NO 628; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activ/inhibit activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 2080 BP; 472 A; 567 C; 570 G; 471 T; 0 other;

Alignment Scores:  
Pred. No.: 1,51e-193 Length: 2080  
Score: 1751.00 Matches: 331  
Percent Similarity: 99.40% Conservative: 0  
Best Local Similarity: 99.40% Mismatches: 1  
Query Match: 98.81% Indels: 1  
DB: 22 Gaps: 0

US-09-848-852A-3 (1-332) x AAI58425 (1-2080)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20  
DB 634 ATGTACTCTGCCACGGTTTGGGGCAATCTGGCCAGAGGGGCGCGAGCTGGC 693  
QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40  
DB 694 ACCCCCTGATGCCAGCATCCTCATCAGCAATGTGTGCAGCATCGGGACCCTGTGGCCCA 753  
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60  
DB 754 GGAGCTTTTTCAGGGCTCAGATTGGGCATGGCAGAAAGGAGGAGCGCTGGGGAGAA 813  
QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80  
DB 814 ACCCGCCAGACAGCCCTCGGAGAGGAGCATGTGACCTCGGTACAGAGCATCTTGA 873  
QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValGluLy 100  
DB 874 CGAATTCTTCAACGATATGGCAGCCTCATACCCCTCAGCACTGATGAGGTAGAGAA 933

QY 100 sLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGl 120  
DB 934 GCTGGAGGACATTTCCAGCAGAGAGTTTCCACCCCTCCAGAGAGGGCTGGTGTGA 993  
QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140  
DB 994 GCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGGTGGAGGGCTCCGAGTGC 1053  
QY 140 aTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160  
DB 1054 TTATAAGCGGCACGTGCTGACCATGGATGGGACCTTGTATGGACAGACAGTGGCT 1113  
QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180  
DB 1114 CAATGACCCAGGTGATGACATGTATGGACACCTGTGTCATGGACACAGTCCCTGAAAAGT 1173  
QY 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLy 200  
DB 1174 GCATTTCTTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGTGA 1233  
QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220  
DB 1234 AAGGTGGACCAAAAACGTGGACATCTCAATAAGGAGCTACTGCTAATCCCATCCACCT 1293  
QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgArgThrIleThrTyrPheAs 240  
DB 1294 GGAGGTGCATTTGGTCCCTCATCTCTGTGATGTGAGGCGACGCCACCATATTTTGA 1353  
QY 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260  
DB 1354 CTCGACGCTACCCCTAAACCGCGCTGCCTAAGCATATTGCCAAGTATCTACAGGCAGA 1413  
QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280  
DB 1414 GCGGTAAAGAAGACCGCATGCTGATTTCCACGAGGCTGGAAGGTACTTCAAAATGAA 1473  
QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300  
DB 1474 TGTGGCCAGGCAGATAATGACAGTGACTGTGTGCTTTTGTGTGTCAGTACTGCAAGCA 1533  
QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgGl 320  
DB 1534 TCTGGCCCTGTCTCAGCCATTCCAGCTTCCACGAGGAGACATGCCCAAACCTTCGTCGGA 1593  
QY 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332  
DB 1594 GATCTACAGGAGCTGTGTCACTGCAAACTCACTGTG 1630  
RESULT 3  
AAF54831  
ID AAF54831 standard; DNA; 2206 BP.  
XX  
AC AAF54831;  
XX  
DT 15-MAY-2001 (first entry)  
XX  
DE Nucleotide sequence of a human SENP2 polypeptide.  
XX  
KW Human; de-sentrinase; SENP; SENP1; SENP2; SENP3; chromosome 12q13.1;  
KW sentrin-specific protease; sentrin; anti-proliferative agent;  
KW anti-viral agent; herpes simplex-1; cytomegalovirus; sentrinisation;  
KW PML; tumour suppressor; acute promyelocytic leukaemia; ss.  
XX  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 52..1758  
FT /\*tag= a  
FT /product= "SENP3"  
XX  
PN WO200109292-A2.  
XX

PD 08-FEB-2001.  
 XX  
 PF 31-JUL-2000; 2000WO-US20884.  
 XX  
 PR 31-JUL-1999; 99US-0146774.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 XX Gong L, Yeh ETH;  
 XX  
 XX WPI; 2001-182947/18.  
 DR P-PSDB; AAB31977.  
 XX  
 XX New polynucleotide encoding de-sentrinase polypeptides, useful as  
 PT antiproliferative or antiviral agents -  
 XX  
 XX Claim 36; Page 120; 122pp; English.  
 XX  
 CC The present sequence encodes a human de-sentrinase (SENP) 3 polypeptide.  
 CC The specification describes SENP1 and SENP2 polypeptides. The SENP1  
 CC gene is found in chromosome 12q13.1. SENP1 is found in the nucleus,  
 CC SENP3 is found in the nucleolus, and SENP3 is found in both locations.  
 CC SENP polypeptides are sentron-specific proteases that remove sentrin  
 CC from some sentrinised peptides, but do not affect proteins modified by  
 CC ubiquitin or NEDD8. SENP polypeptides are used to identify specific  
 CC modulators of SENP. These modulators are potential anti-proliferative  
 CC and anti-viral agents such as herpes simplex-1 or cytomegalovirus.  
 CC SENP polypeptides are used for studying the role of sentrinisation in  
 CC the biological function of PML, a tumour suppressor implicated in  
 CC development of acute promyelocytic leukaemia. Fragments of SENP  
 CC polynucleotides are used as hybridisation probes and amplification  
 CC primers for detecting gene expression or preparing mutated sequences,  
 CC also as antisense sequences for inhibiting SENP expression.  
 XX  
 SQ Sequence 2206 BP; 491 A; 612 C; 605 G; 498 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.64e-193 Length: 2206  
 Score: 1751.00 Matches: 331  
 Percent Similarity: 99.40% Conservative: 0  
 Best Local Similarity: 99.40% Mismatches: 1  
 Query Match: 98.81% Indels: 1  
 DB: 22 Gaps: 0  
 US-09-848-852A-3 (1-332) x AAP54831 (1-2206)  
 QY 1 MetTyrSerAlaGlnArgPheThrGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20  
 DB 759 ATGTACTCTGCCCAACGGTTTGGGGGACAAATCTGGGCCAGAGGGGCGCAGCTTGGC 818  
 QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG1 40  
 DB 819 ACCCCCTGATGCCAGCATCTCATCAGCATGTGTCAGCATCGGGGACCATGTGGCCCA 878  
 QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60  
 DB 879 GGAGCTTTTTCAGGGCTTCAGATTGGCATGGCAGAGAGAGAGAGAGAGAGAGAGAG 938  
 QY 60 sAlaGlyGlnHisSerProLeuArgGluHisValThrCysValGlnSerIleLeuAs 80  
 DB 939 AGCCGGCCACAGACCCCTCGCAGAGGAGCATGTGACCTGGCTGACAGACATCTTGA 998  
 QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLy 100  
 DB 999 CGAATTCCTTCAACGTATGCGAGCCTCATACCCCTCAGCAGCTGATGAGTAGAGAA 1058  
 QY 100 LeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuG1 120  
 DB 1059 GCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGGCCTGGTGTGCA 1118  
 QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140  
 DB 1119 GCTGATCCAGCTTTACACAGCGGATGTCAGGCAATGCCAATGCTGGTGGGGGCTTCGAGTGGC 1178

140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160  
 1179 TTATAGCGGCACGCTGCTGACCATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCT 1238  
 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180  
 1239 CAATGACCAAGGTGATGAACATGTATGAGACCTGGTGCATGGACACAGTCCCTGAAAGGT 1298  
 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLy 200  
 1299 GCATTTCTTCAATAGTTTCTTCTATGATAAATCCGTACCAAGGGTATGATGGGTGAA 1358  
 200 sArgTyrThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220  
 1359 AAGTGGACCAAAAACCTGGACATCTTCAATAAGGAGCTACTGCTAATCCCCATCCACCT 1418  
 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240  
 1419 GGAGGTGCATTGGTCCCTCATCTCTGTGATGTGAGCGACGACCATCATCTATTTTGA 1478  
 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaG1 260  
 1479 CTCGCAGCGTACCTAAACCGCGCTGCCCTAAGCATATTTGCCAAGTATCTACAGGCAGA 1538  
 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280  
 1539 GCGGGTAAGAAAGACCGACTGGATTTCCACAGGCGTGGAAAGGTACTTCAAAATGAA 1598  
 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysH1 300  
 1599 TGTGGCCAGGCAGCAATAATGACAGTGTGCTGTGCTTTGTTGTCAGTACTGCAAGCA 1658  
 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgG1 320  
 1659 TCTGGCCCTGTCTCAGCCATTTCAGCTTACCCAGAGGACATGCCAAACTTCGTCGGCA 1718  
 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332  
 1719 GATCTACAGGAGCTGTGTCAGTGCAGCAACTCAGCTGTG 1755  
 RESULT 4  
 AAD05541/c  
 ID AAD05541 standard; cDNA; 1507 BP.  
 XX  
 AC AAD05541;  
 XX  
 DT 18-JUL-2001 (first entry)  
 XX  
 DE Human secreted protein-encoding gene 11 cDNA clone HOF0B27, SEQ ID NO:60.  
 XX  
 KW Human: secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 KW endocrine disorder; infection; wound healing; vulnerability;  
 KW cell culture; chemotaxis; food additive;  
 KW binding partner identification; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 CDS 416..715  
 FT /\*tag= a  
 FT /product= "Human secreted protein precursor"  
 FT /note= "Does not include start codon"  
 FT /partial  
 FT sig\_peptide 416..493  
 FT /\*tag= b

```

FT mat_peptide 494..712
FT /*tag= c
FT /product= "Mature human secreted protein"
PN WO200134767-A2.
XX 17-MAY-2001.
XX 01-NOV-2000; 2000WO-US30036.
XX 05-NOV-1999; 99US-0163576.
XX 27-JUL-2000; 2000US-0221366.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Soppet DR, Komatsoulis G, Shi Y, Olsen HS, Ruben SM;
XX WPI; 2001-316492/33.
XX P-PSDB; AAE01721.
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX Claim 1; Page 447; 540pp; English.
XX AAD05492-AAD05564 represent cDNAs corresponding to 22 human secreted
XX protein genes, and AAE01672-AAE01743 represent the proteins they encode.
XX AAE01744-AAE01763 represent human secreted protein fragments or variants.
XX The secreted proteins and their genes are useful for preventing,
XX treating or ameliorating medical conditions, e.g., by protein or gene
XX therapy. Pathological conditions can be diagnosed by determining the
XX amount of the new protein in a sample or by determining the presence of
XX mutations in the new genes. Specific uses are described for each of the
XX 22 genes, based on the tissues in which they are most highly expressed,
XX and include developing products for the diagnosis or treatment of
XX proliferative disorders, cancer, tumours, foetal and developmental
XX abnormalities, haematopoietic disorders, diseases of the immune system,
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX allergies, neurological disorders (e.g., Alzheimer's disease,
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma,
XX skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX cardiovascular disorders, angioecic disorders, kidney disorders,
XX gastrointestinal disorders, pregnancy-related disorders, endocrine
XX disorders, and infections. The proteins can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues, to identify their
XX cognate ligands or binding partners, and in chemotaxis, and can be used
XX as a food additive or preservative to modify storage properties.
XX Antibodies specific for a protein of the invention can be used in
XX alleviating symptoms associated with the disorders mentioned above, and
XX in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
XX immunosorbent assay (ELISA). The present sequence represents a human
XX secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 1507 BP; 357 A; 385 C; 354 G; 401 T; 10 other;

Alignment Scores:
Pred. No.: 3,68e-188 Length: 1507
Score: 1703.00 Matches: 320
Percent Similarity: 99.69% Conservative: 1
Best Local Similarity: 99.38% Mismatches: 1
Query Match: 96.11% Indels: 0
DB: 22 Gaps: 0

US-09-848-852A-3 (1-332) x AAD05541 (1-1507)

Qy 11 IleTrpAlaArgGlyAlaHisLeuAlaProAspAlaSerIleLeuIleSerAsn 30
Db 1453 CTTTGGCCAGAGGGAGCGCANTTGGCACCCTTGATGCCAGCATCTCTATCAGCAAT 1394
Qy 31 ValCysSerIleGlyAspHisValAlaGlnGlnLeuLeuPheGlnGlySerAspLeuGlyMet 50
Db 1453 CTTTGGCCAGAGGGAGCGCANTTGGCACCCTTGATGCCAGCATCTCTATCAGCAAT 1394

```

```

Db 1393 GTGTGCAGCATCGGGAGCACCATGTGCCAGGAGCTTTTTCAGGGCTCAGATTTGGGCATG 1334
Qy 51 AlaGluGluAlaGluArgProGlyGluLysAlaGlyGlnHisSerProLeuArgGluGlu 70
Db 1333 GCAGAAGAGGCAGAGAGCGCTGGGAGAAAGCCGCCAGCACAGCCCCCTGCGAGAGGAG 1274
Qy 71 HisValThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGlySerLeuIle 90
Db 1273 CATGTGACCTGGTACAGAGCATCTTGGACGAATTCCTTCAACAGCTATGGCAGCCTCAT 1214
Qy 91 ProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGlnIlePheSer 110
Db 1213 CCCCTCAGCACTGAYGAGGTAGTAGAARTRGAGGACATTTTCAGCAGAGATTTCY 1154
Qy 111 ThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMetProGly 130
Db 1153 ACMCTTCCAGGAAGGGCTGGTGTGAGCTGATCCAGTCTACAGCGGATGCCAGGC 1094
Qy 131 AsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThrMetAspAsp 150
Db 1093 AATGCCATGGTGAGGGCTTCCGAGTGGCTTATAAGCGGCACGTGCTGACCATGGATGAC 1034
Qy 151 LeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrGlyAsp 170
Db 1033 TTGGGGACCTTGTATGGACAGAACTGGCTCAATGACCAAGGTGATGAACATGTATGGAGAC 974
Qy 171 LeuValMetAspThrValProGluLysValHisPheAsnSerPhePheTyrAspLys 190
Db 973 CTGGTCATGGACACAGTCCCTGAAAGGTGCATTTCTCAATAGTTTCTTCATGATAAA 914
Qy 191 LeuArgThrLysGlyTyrAspGlyValLysArgTyrThrLysAsnValAspIlePheAsn 210
Db 913 CTCCTGATCCAGGGTTATGATGGGGTGAAGAGTGGACCAAAAACGTGGACATCTCAAT 854
Qy 211 LysGluLeuLeuLeuIleProIleHisLeuGluValHisTrpSerLeuIleSerValAsp 230
Db 853 AAGGAGCTACTGCTAATCCCATCCACCTGGAGGTGCATTTGGTCCCTCATCTCTGTGAT 794
Qy 231 ValArgArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgArgCysPro 250
Db 793 GTGAGCGCAGCGCACCCTACCTATTTTACTCGCAGCGTACCCTAAACCGCGCTGCCCT 734
Qy 251 LysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHis 270
Db 733 AAGCATATGGCAAGTACTACAGCAGAGCGGTAAAGAAAGACCGAGTGGATTTCCAC 674
Qy 271 GlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAsnAspSerAspCys 290
Db 673 CAGGGCTGGAAAGGTACTTCAAAATGAATGTGGCCAGGCAGAGTAATGACAGTACTGT 614
Qy 291 GlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerGlnPropheSerPheThr 310
Db 613 GGTGCTTTTGTGTTGTCAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCAGCTTACC 554
Qy 311 GlnGlnAspMetProLysLeuArgGlnIleTyrLysGluLeuLeuCysHisCysLysLeu 330
Db 553 CAGCAGGACATGCCCAAACTTCGTCCGCAGATCTACAGGAGCTGTGTCTACATGCAAACTC 494
Qy 331 ThrVal 332
Db 493 ACTGTG 488
RESULT 5
AAD05502
XX AAD05502 standard; cDNA; 1934 BP.
XX AC AAD05502;
XX XX
XX 18-JUL-2001 (first entry)
XX DE Human secreted protein-encoding gene 11 cDNA clone HOF0B27, SEQ ID NO:21.
XX KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;

```

foetal abnormality; developmental abnormality; haematopoietic disorder; immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis; Parkinson's disease; cognitive disorder; schizophrenia; skin disorder; psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder; inflammation; neurological disorder; Alzheimer's disease; food additive; angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy; pregnancy-related disorder; endocrine disorder; infection; wound healing; cell culture; chemotaxis; vulvurary; binding partner identification; gene therapy; ss.

OS Homo sapiens.

Key Location/Qualifiers  
 CDS 441..713  
 /\*tag= a  
 /product= "Human secreted protein precursor"  
 sig\_peptide 441..521  
 /\*tag= b  
 mat\_peptide 522..710  
 /\*tag= c  
 /product= "Mature human secreted protein"

WO200134767-A2.

17-MAY-2001.

01-NOV-2000; 2000WO-US30036.

05-NOV-1999; 99US-0163576.

27-JUL-2000; 2000US-0221366.

(HUMA-) HUMAN GENOME SCI INC.

Soppet DR, Komatsoulis G, Shi Y, Olsen HS, Ruben SM;

P-PSDB; AA01682.

Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition -

Claim 1; Page 417-418; 540pp; English.

AA05492-AA05564 represent cDNAs corresponding to 22 human secreted protein genes, and AA01672-AA01743 represent the proteins they encode. AA01744-AA01763 represent human secreted protein fragments or variants. The secreted proteins and their genes are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 22 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infectious. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.

SQ Sequence 1934 BP; 423 A; 542 C; 551 G; 418 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 6,94e-188 Length: 1934  
 Score: 1702.00 Matches: 324  
 Percent Similarity: 97.90% Conservative: 2  
 Best Local Similarity: 97.30% Mismatches: 6  
 Query Match: 96.05% Indels: 1  
 DB: 22 Gaps: 0  
 US-09-848-852a-3 (1-332) x AAD05502 (1-1934)  
 QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAla-HisLeuAl 20  
 DB 496 CTGTAATCTTCCCAATGGCTTTGGGGGACCCCGGGCAGAGAGGGGGTCTGGC 555  
 QY 20 aProProAspAlaSerIleLeuLeuSerAsnValCysSerIleGlyAspHisValAlaGl 40  
 DB 556 ACCCCCTGATGCCAGCATCTCATCAGCAATGTGTGCAGCATCGGGACCATGTGCCCA 615  
 QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60  
 DB 616 GGAGCTATTTTCAGGGCTCAGATCTGGGACCCGAGAGAGGAGGCGGGGAGAA 675  
 QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80  
 DB 676 AGCGGCCAGCAGACGCCCTCGGGGAGGAGCATGTGACCTGGTGCAGAGCATCTTGA 735  
 QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLy 100  
 DB 736 TGAATTCCTTCAAACTTATGGCAGCTCATCCGCCCTCAGCACGAGGAGGTAGAGAA 795  
 QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGl 120  
 DB 796 ATTAGAGAGCATTTTCCAGGAGGAGTTCTACACCTTCAGAGAGGGCTGTGTGTGCA 855  
 QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140  
 DB 856 GCTGATCCAGTCGTACCAGCGGATGCCAGCAACGCCATGGTGAGGGGCTTCGGGTGGC 915  
 QY 140 aTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160  
 DB 916 CTACAAGCGCAGCTGCTGACCATGTGATGACCTGGGCACCTTATATGGACAGAACTGGCT 975  
 QY 160 uasnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGlyLysVa 180  
 DB 976 CAACGACCCAGGTGAAGATGTACGAGACCTGGTGTACGACCCGCTCCCGGAGAGGT 1035  
 QY 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200  
 DB 1036 GCATTTCTTCACAGTTTCTTCTACGATAAGCTCCCTACCAAGGGTTACGATGAGTGA 1095  
 QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220  
 DB 1096 AAGGTGGACCAAAACGTGGACATCTTCAATAGGAGCTCCTGCTAATCCCATCCACCT 1155  
 QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240  
 DB 1156 GGAGGTGCACCTGGTCCCTCATCTCTTACGTGAGCGCGCGCACCATCACGTATTTCGA 1215  
 QY 240 pSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260  
 DB 1216 CTCGACGCCACCCCTAAACCGCGCTGCCCTAAGCATATTTGCCAAGTATCTACAGCAGA 1275  
 QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280  
 DB 1276 GGCAGTGAAGAAAGACCGCTGGATTTCCACCAGGGCTGAAAGGTTATTTCAAAATGAA 1335  
 QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300  
 DB 1336 TGTGGCGAGGAGAAATATGACAGTGACTGCGGTGCTTTGTGTACACTACTGCAAGCA 1395  
 QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgGl 320

Db 1396 CCTGGCCCTGTCTCAGCCATTGAGTTCACCCAGCAGGACATGCCCAAACTTCGCGCGCA 1455  
|||||  
QY 320 nileTyrLysGluLeuCysHisCysLysLeuThrVal 332  
|||||  
Db 1456 GATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTG 1492  
|||||  
RESULT 6  
AAI60211/c  
ID AAI60211 standard; cDNA; 1804 BP.  
XX AAI60211;  
AC AAI60211;  
XX AAI60211;  
DT 22-OCT-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 4200.  
DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Dermanac RT;  
XX  
DR WPI; 2001-442253/47.  
XX  
XX P-PSDB; AAM41055.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX Claim 1; SEQ ID NO 4200; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities, chemotactic/chemokinetic activity, haemostatic  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
XX specification.  
XX  
XX Sequence 1804 BP; 431 A; 477 C; 468 G; 428 T; 0 other;

Alignment Scores: 1.4e-187 Length: 1804  
Pred. No.: 1699.00 Matches: 331  
Score: 98.22% Conservative: 0  
Percent Similarity: 98.22% Mismatches: 1  
Best Local Similarity: 95.88% Indels: 5  
Query Match: 22 Gaps: 0  
DB: 22  
US-09-848-852A-3 (1-332) x AAI60211 (1-1804)  
QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20  
|||||  
Db 1485 ATGTACTCTGCCCAACGGTTTTTGGGGGACAATCTGGGCCAGAAAGCGGAGCTTGGC 1426  
|||||  
QY 20 aProPaspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40  
|||||  
Db 1425 ACCCCCTGATGCAGCATCTCATCAGCAATGTGTGACATCGGGGACCATGTGGCCCA 1366  
|||||  
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGlu-L 60  
|||||  
Db 1365 GGAGCTTTTTCAGGCTCAGATTTGGGCATGCGCAGAAAGAGGAGAGCCCTGGGGAGAA 1306  
|||||  
QY 60 ysAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuA 80  
|||||  
Db 1305 AAGCCGGCCAGCACAGCCCTCGGAGAGGAGCATGTGACCTGCGTACAGAGCATCTGG 1246  
|||||  
QY 80 spGluPheLeuGlnThr-TyrGlySerLeuIleProLeuSerThrAspGluValValGlu 99  
|||||  
Db 1245 ACGAATTCCTTCAACGGTATGGCAGCCCTCATACCCCTCAGCACTGATGAGTAGTAG 1186  
|||||  
QY 100 LysLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeu 119  
|||||  
Db 1185 AAGCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGCCCTGGTGTG 1126  
|||||  
QY 120 GlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgVal 139  
|||||  
Db 1125 CAGCTGATCCAGTCTTACCAGCGGATGCCAGCAATGCCATGGTAGGGGCTTCCGAGTG 1066  
|||||  
QY 140 AlaTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrp 159  
|||||  
Db 1065 GCTTATAAGCGGCAGCTGTGTGACCATGTGATGACTTTGGGGACCTGTGATGACAGAACTGG 1006  
|||||  
QY 160 LeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGlu-Ly 179  
|||||  
Db 1005 CTCATATGACCGTATGATGACATGTATGAGACCTGCTGATGACACACATCCCTGAAAAA 946  
|||||  
QY 179 sValHisPheAsnSerPhePheTyr-AspLysLeuArgThrLysGlyTyrAspGlyY 199  
|||||  
Db 945 GGTGCATTTCTCAATAGTTTCTCTATTTGATAAACTCCGTACCAAGGTTTATGATGGG 886  
|||||  
QY 199 aLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleH 219  
|||||  
Db 885 TGAAGAGGTGGACAAAAACGTGGACATCTTCAATAAGAGGACTACTGCTTAATCCCATCC 826  
|||||  
QY 219 isLeuGluValHisTrpSerLeuIleSerValAspValArgArgArgThrIleThrTrp 239  
|||||  
Db 825 ACCTGGAGTGCATTTGGTCCCTCATCTCTGTTGATGTGAGGCGACGCCACCATCATTT 766  
|||||  
QY 239 heAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnA 259  
|||||  
Db 765 TTGACTCGCAGCGTACCCTTAACCCGCGCTAAAGCATATTGTCAGATATTCACAGG 706  
|||||  
QY 259 isGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysM 279  
|||||  
Db 705 CAGAGCGGTGAAGAAAGACCGACTGGATTTCCACAGGCGCTGGAAGGTTACTTCAAAA 646  
|||||  
QY 279 etAsnValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysL 299  
|||||  
Db 645 TGAATGTGGCCAGGAGCAATAATGACAGTGTGCTTTTGTGTGTGAGTACTGCA 586  
|||||  
QY 299 ysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgA 319  
|||||

Db	585	AGCATCTGGCCCTGTCTCAGCCATTCAGCTTACCACGAGGACATGCCAACTTGCTC	526
Qy	319	rgGlnIleTyLyrsGluLeuCysHisCysLysLeuThrVal	332
Db	525	GGCAGATCTACAAGGAGCTGTGTCTACTGCAAACTCACTGTG	485
RESULT 7			
AAI60212/c			
ID	AAI60212	standard; cDNA; 1804 BP.	
XX	AC	AAI60212;	
XX	DT	22-OCT-2001 (first entry)	
XX	XX	Human polynucleotide SEQ ID NO 4201.	
XX	XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;	
KW	KW	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	KW	leukaemia; ss.	
OS	OS	Homo sapiens.	
XX	XX	WO200153312-A1.	
XX	PN	26-JUL-2001.	
PD	PD	26-DEC-2000; 2000WO-US34263.	
XX	XX	21-JAN-2000; 2000US-0488725.	
PR	PR	25-APR-2000; 2000US-0552317.	
PR	PR	09-JUL-2000; 2000US-0598042.	
PR	PR	19-JUL-2000; 2000US-0620312.	
PR	PR	03-AUG-2000; 2000US-0653450.	
PR	PR	14-SEP-2000; 2000US-0662191.	
PR	PR	19-OCT-2000; 2000US-0693036.	
PR	PR	29-NOV-2000; 2000US-0727344.	
XX	XX	(HYSE-) HYSEQ INC.	
XX	XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
XX	XX	WPI; 2001-442253/47.	
XX	XX	P-PSDB; AAM41056.	
XX	XX	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	PT	such as central nervous system injuries -	
XX	XX	Claim 1; SEQ ID NO 4201; 10078pp; English.	
PS	PS	The invention relates to human nucleic acids (AAI57798-AAI61369) and	
CC	CC	the encoded polypeptides (AAM38642-AAAM42213) with nootropic,	
CC	CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	CC	central nervous system diseases, such as	
CC	CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	CC	utilisation of the activities such as: immune system suppression,	
CC	CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	CC	assays for receptor activity, arthritis and inflammation, leukaemias and	
CC	CC	C.N.S disorders.	
CC	CC	Note: The sequence data for this patent did not form part of the printed	
XX	XX	specification.	
XX	XX	Sequence 1804 BP; 431 A; 477 C; 468 G; 428 T; 0 other;	

QY 319 rgGlnIleTyrLysGluLeuCysHisCysLysLeuThrVal 332  
ID ABQ54216 standard; cDNA; 1933 BP.  
XX AC ABQ54216;  
XX DT 22-AUG-2002 (first entry)  
XX DE Human ovarian antigen HOF0B27 cDNA, SEQ ID NO:96.  
XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KW infertility; pregnancy disorder; anovulation; reproductive system disorder;  
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; polycystic ovary syndrome;  
KW inflammatory condition; immune disorder; blood disorder;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disorder; urinary system disorder; drug screening;  
KW gene therapy; chromosome mapping; forensic analysis;  
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KW antiinflammatory; gynaecological; reproductive; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200200677-A1.  
XX PD 03-JAN-2002.  
XX PR 07-JUN-2001; 2001WO-US18569.  
XX PR 07-JUN-2000; 2000US-209467P.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Birse CE, Rosen CA;  
XX WPI: 2002-147878/19.  
XX P-FSDB: ABP41139.  
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
PT ovarian cancer), immune disorders, cardiovascular disorders and  
PT neurological diseases -  
XX Claim 1; SEQ ID No 96; 2922pp; English.  
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
XX ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
XX encompasses polypeptides 90% identical and polynucleotides 95% identical  
XX to the sequences of the invention. The invention additionally relates to  
XX recombinant vectors and host cells comprising human ovarian antigen  
XX polynucleotides, antibodies against human ovarian antigens, and the use  
XX of ovarian antigen polynucleotides and polypeptides in diagnosing,  
XX treating, prognosing or preventing various ovary and/or breast-related  
XX disorders. Such conditions include ovarian cancer and breast cancer, and  
XX metastatic tumours of ovarian or breast origin, reproductive system  
XX disorders (e.g., infertility, disorders of pregnancy, anovulation,  
XX polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
XX disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
XX shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
XX vaginitis), immune disorders (e.g., congenital and acquired  
XX immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
XX blood-related disorders (e.g., anaemia), cardiovascular disorders,  
XX respiratory disorders, neurological disorders, gastrointestinal disorders  
XX and urinary system disorders. Ovarian antigen polypeptides and  
XX polynucleotides may also be used in screening for compounds which  
XX modulate ovarian antigen expression or activity. The polynucleotides may  
XX further be used for gene therapy, chromosome mapping, in the  
XX identification of individuals and in forensic analysis, and the

CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents cDNA encoding a human ovarian antigen of the  
CC invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 1933 BP; 423 A; 541 C; 551 G; 418 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 1.27e-184 Length: 1933  
Score: 1674.00 Matches: 323  
Percent Similarity: 97.60% Conservatives: 2  
Best Local Similarity: 97.00% Mismatches: 7  
Query Match: 94.47% Indels: 2  
DB: 24 Gaps: 0  
  
US-09-848-852A-3 (1-332) x ABQ54216 (1-1933)  
QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTTPAlaAtgArgGlyAla-HisIleuAl 20  
DB 496 CTGTACTCTTCCCAATGGCTTTGGGGACCCCGGGCCAGAGGGGAGCGGTCTGC 555  
QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40  
DB 556 ACCCCCTGATGCCAGCATCTCATCAGCAATGTGTGACGATCGGGGACCATGTGCCCA 615  
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60  
DB 616 GGAGCTATTTCAGGGCTCAGATCTGGGCACCCAGAGGAGGAGCGCGGGGAG-AA 674  
QY 60 sAlaGlyGlnHisSerProLeuArgGluHisValThrCysValGlnSerIleLeuAs 80  
DB 675 AGCCGGCCAGCACACCCCTCGGGAGAGCATGTGACCTCGTGACAGAGCATTTTGA 734  
QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValuLy 100  
DB 735 TGAATTCCTTCAAACTTATGGCAGCCTCATCCCTCAGCACGGAGGAGTAGAGAA 794  
QY 100 sLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValuLeuGl 120  
DB 795 ATTAGAGGACATTTTCCAGCAGGAGTTCTCTACACTTCCAGGAAGGGCTGTGCTGCA 854  
QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140  
DB 855 GCTGATTCAGTCTGTACCGGATGCCAGGCAACGCCATGGTGGGGGCTTCCGGGTGC 914  
QY 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160  
DB 915 CTACAAGCGGCACGTGCTGACCATGGATGACCTGGGACCTTATATGGACAGAACTGGCT 974  
QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180  
DB 975 CAACGACGAGGTGATGAACATGTAGGAGACCTGTGTCATGGACAGCGTCCCGGAGAA 1034  
QY 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVal 200  
DB 1035 GCATTTCCTTCAACAGTTTCTTCTAGATAAGCTCCGTACCAAGGGGTAGTAGGAGTAA 1094  
QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220  
DB 1095 AAGGTGGACCAAAACGTGGACATCTTCAATAAGGAGCTCCTGCTTAATCCCATCCACCT 1154  
QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrPheAs 240  
DB 1155 GGAGGTGCACTGGTCCCTCATCTCTGTGACGTGAGGCGGCGCACCATCACGTATTTCGA 1214  
QY 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260  
DB 1215 CTCGAGCGGACCCCTAAACCGCGCTGCCCTAAGCATATTCGCAAGTATCTACAGGCGA 1274  
QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280

|||||  
Db 1275 GGCAGTGAAGAAAGCCGGCTGGATTTCACAGGGCTGGAAGGTTATTTCAAAATGAA 1334  
Qy 280 nValAlaArgGlnAsnAsnSerAspCysGlyAlaPheValLeuGlnTyrCysLysHI 300  
Db 1335 TGTGGCCAGGAGCAATAATGACAGTGACTGCGGTGCTTTGTGTACAGTACTGCAAGCA 1394  
Qy 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgGI 320  
Db 1395 CTTGGCCCTGTCACGCCATTACGCTTCACCCAGCAGGACATGCCCAAACTTCGCCGGCA 1454  
Qy 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332  
Db 1455 GATCTACAAGGAGGTGTGTCACTGCAAACTCACTGTG 1491  
RESULT 9  
AA158426  
ID AA158426 standard; cDNA; 2029 BP.  
XX  
AC AA158426;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 629.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
PN WC200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
{HYSE-} HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
DR P-PSDB; AAM39270.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Claim 1; SEQ ID NO 629; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
..

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 2029 BP; 460 A; 560 C; 553 G; 456 T; 0 other;  
Alignment Scores:  
Pred. No.: 6,39e-181 Length: 2029  
Score: 1642.50 Matches: 314  
Percent Similarity: 94.29% Conservative: 0  
Best Local Similarity: 94.29% Mismatches: 1  
Query Match: 92.69% Indels: 18  
DB: 22 Gaps: 1  
US-09-848-852A-3 (1-332) x AA158426 (1-2029)  
Qy 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAlaHis-LeuAl 20  
Db 634 ATGTACTCTGCCCAACGGTTTGGGGGACAAATCTGGGCCAGAGGGGAGCGAGCTTGGC 693  
Qy 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGI 40  
Db 694 ACCCCCTGATGCCAGCATCCTCATCAGCAATGTGTGCAGCATCGGGGACCATGTGCCCA 753  
Qy 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60  
Db 754 GGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGGAGGAGGCGCTGGGAGAA 813  
Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80  
Db 814 AGCCGCCACAGCAGCCCCCTGGGAGAGGAGCATGTGACCTGGGTACAGAGCATCTTGA 873  
Qy 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLy 100  
Db 874 CGAATTCTCTCAAACGTATGGCAGCGCTCATACCCCTCAGCACTGATGAGGTAGTAGAA 933  
Qy 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGI 120  
Db 934 GCTGGAGGACATTTTCCAGCAGAGAGTTTCCACCCCTTCCAGGAAGGGCTGGTGTGCA 993  
Qy 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140  
Db 994 GCTGATCCAGTCTTACCACCGGATGCCAGCAATGCCATGCTGAGGGGCTTCGAGTGGC 1053  
Qy 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160  
Db 1054 TTATAAGCGGCACGTGCTGACCATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCT 1113  
Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180  
Db 1114 CAATGACCAAGGTGATGAACATGTATGGAGACCTGTGTCATGGACAGACAGTCCCTGAAAAGGT 1173  
Qy 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200  
Db 1174 GCATTTCTTCATAGTTTCTTCTATGATGATAAACTCCGTACCAAGGGTTATGATGGGGTGA 1233  
Qy 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220  
Db 1234 AAGGTGGACCAAAACGTGGACATCTTCAATAAGGAGCTACTGCTAAATCCCATCCACCT 1293  
Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgArgThrIleThrTyrPheAs 240  
Db 1294 GGAGGTGATTCGTCCTCTCATCTCTGTGTGATGTGAGGGGACGACCATCACCATTATTTGA 1353  
Qy 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGI 260  
Db 1354 CTCGCGAGCGTACCCTAAACCGCGCTGCCCTAAGCATATTTGCCAAGTATCTACAGGCAGA 1413  
Qy 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280  
|||||



Db 1414 GCGGTAAGAAAGACCGACTGGATTTCACACAGGGCTGGAAGGTTACTTCAAAATG-- 1471  
QY 280 nValAlaArgGlnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysIshi 300  
Db 1472 -----TACTGCAAGCA 1482  
QY 300 sleuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgG1 320  
Db 1483 TCTGGCCCTGTCTCAGCCATTTCAGCTTCACCCAGCAGGACATGCCCAAACITTCGTCGGCA 1542  
QY 320 nileTyrLysGluLeuCysHisCysLysLeuThrVal 332  
Db 1543 GATCTACAAAGGAGCTGTCTCACTGCAAACTCACTGTG 1579  
RESULT 10  
AAH13708  
ID AAH13708 standard; cDNA; 1578 BP.  
XX AAH13708;  
AC AAH13708;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:10590.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 10590; 2537pp + CD ROM; English.  
XX  
PS  
CC The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 1578 BP; 339 A; 372 C; 439 G; 428 T; 0 other;  
Alignment Scores:  
Pred. No.: 3.07e-85 Length: 1578  
Score: 820.00 Matches: 186  
Percent Similarity: 50.00% Conservatives: 10  
Best Local Similarity: 47.45% Mismatches: 17  
Query Match: 46.28% Indels: 179  
DB: 22 Gaps: 4  
US-09-848-852A-3 (1-332) X AAH13708 (1-1578)  
QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20  
Db 274 ATGTACTCTGCCCAACGGTTTTTGGGGGACAATCTCTGGGCAGAAAGGAGCGAGCTTGGC 333  
QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG1 40  
Db 334 ACCCCCTGATGCCAGCATCTCATCAGCAATGTGTGACATCGGGGACCATGTGGCCCA 393  
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLyl 60  
Db 394 GGAGCTTTTTCAGGCTCAGATTTGGGCATGCCAGAGGAGAGAGGCTTGGGAGAA 453  
QY 60 sAlaGlyGlnHisSerProLeuArgGluHisValThrCysValGlnSerIleLeuAs 80  
Db 454 AGCCGCCAGCAGACAGCCCTCGAGAGAGCATGTGACCTCGGTACAGAGCATCTTGA 513  
QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValAlaGluLyl 100  
Db 514 CGAATTCCTTCAACGATATGGCAGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGAA 573  
QY 100 sleuGluAspIlePheGlnGlnPheSerThrProSer----- 113  
Db 574 GCTGGAGGACATTTTCCAGCAGGAGTTCACACCCCTTCCAGGTGAGGCTTGAAGCCCT 633  
QY 113 ----- 113  
Db 634 CCTTGAAGAGGGCTGGGGCTTGGGGATGTGGAGAGAATACTGCTGCTTTTCTTCCA 693  
QY 113 ----- 113  
Db 694 TAGGCTCTAGTTGGGGAGAGAGAGCTAGAGCTGAAGGGGAGAGACTCTGACGAGGCTG 753  
QY 113 ----- 113  
Db 754 CCAATCTTGGAAAGCTGATGGGAGAGTCTTTACCTGGGACCTGAAATGTTCTACCTGA 813  
QY 113 ----- 113  
Db 814 GTAGTCATGTTTATCTTCTTGGGGAGTGGGCTTTTCGAGGTCTCTCAGAAGAGCACCATCA 873  
QY 113 ----- 113  
Db 874 TGAGCCAGAAAAAGGAGGATCAGTTAGAATTTGTATTCAGGGAGTAGTAGTATTTCTG 933  
QY 114 -----ArgLysGlyLe 117  
Db 934 TGTGCCCCAGCTGCATCATCTTTTGTGTGACTCCACCCTTGGCCCTACTCAGGAAGGGCCT 993  
QY 117 uValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPh 137  
Db 994 GGTGTTGCAGCTGATCCAGTCTTACCAGGGATGCCAGCAATGCATGCTGAGGGGCTT 1053  
QY 137 eArgValAlaTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyG1 157  
Db 1054 CCGATGGGCTTATAAGCGCAGCTGCTGACCATGGATGATCTTGGGGACCTTGTATGGACA 1113  
QY 157 nAsnTrpLeuAsnAsp----- 162



KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder; ss.

OS Homo sapiens.

XX WO200153455-A2.

PN 26-JUL-2001.

PD 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

XX WPI, 2001-457603/49.

DR P-P5DB; AAM25617.

Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX Claim 1; Page 494; 1217pp; English.

CC AAM99166 to AAM99904 encode the human proteins given in AAM25225 to  
 CC AAM25963. The proteins can have activities based on the tissues and  
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;  
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,  
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders.

XX Sequence 1938 BP; 565 A; 366 C; 425 G; 581 T; 1 other;

Alignment Scores:

Pred. No.: 1.36e-77 Length: 1938  
 Score: 755.50 Matches: 148  
 Percent Similarity: 73.19% Conservatives: 54  
 Best Local Similarity: 53.62% Mismatches: 65  
 Query Match: 42.64% Indels: 9  
 DB: 22 Gaps: 4

US-09-848-852A-3 (1-332) x AAM99558 (1-1938)

QY 58 GlyGluLysAlaGlyClnHisSerProLeuArgGluGluHisValThr---CysValGln 76

Db 3 GGAAGACACAGTCAGAAAGCCCTCTCTGGATGATGAACAGCTGTCAGTCTGCTTCT 62

QY 77 SerIleLeuAspGluPheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGlu 96

Db 63 GGATTCCTAGATGAGGTATATGAAGAGTATGGCAGTGTGGTCCACTCAGTGAAGAA 122

QY 97 ValValGluLysLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGly 116

Db 123 GTCTTGAAGATTAAGATGCTTTAATGAAGACTTTTCT-----AATAGAAACCA 176  
 QY 117 LeuValLeuGlnLeuIleGlnSerTyrglnArgMetProGlyAsnAlaMetValargGly 136  
 Db 177 TTTATCAATAGGGAATAACAACTATCGGCCACAGCATCAAAATGT-----AAC 227  
 QY 137 PheArgValAlaIleTyrglyArgHisValLeuThrMetAspLeuGlyThrLeuTyrgly 156  
 Db 228 TTCCGTATCTTCTATAATAAACACATGCTGGATATGGAGCAGCTCTGGATGT 287  
 QY 157 GlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThrVal 176  
 Db 288 CAGAACTGGCTCAATGACCCAGCTCATTAATATGATGTGAGCTGATATGATGATGATG 347  
 QY 177 ProGluLysValHisPhePheAsnSerPhePheTyrglyAspLeuArgThrLysGlyTy 196  
 Db 348 CCAGACAAAGTTCACTTCTTCAACAGCTTTTTCATAGACAGCTGTAACCAAGGATAT 407  
 QY 197 AspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeu 216  
 Db 408 AATGGAGTAAAGATGGACTAAAGGGTGGATTTGTTTAAAGAGAGTCTTCTGTGAT 467  
 QY 217 ProIleHisLeuGluValHisTrpSerLeuIleSerValAspValargArgThrIle 236  
 Db 468 CCTATTCACTGGAAGTCCACTGGTCTCTCATTAATGTCACACTCTCTAATCGAATATT 527  
 QY 237 ThrTyrglyAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTy 256  
 Db 528 TCATTTTATGATCCCAAGGCATTCAATTTAAGTTTGTGTAGAGAATAAAGAAGTAT 587  
 QY 257 LeuGlnAlaGluAlaValLysLysAspArg-LeuAspPheHisGlnGlyTrpLysGlyTy 276  
 Db 588 TTGCTGACTGAAGCCAGAGAAATAATAGACTGTAATCTT---CAGGGTTGCAGACTGC 644  
 QY 276 rPheLysMetAsnValAlaArgGlnAsnAspSerAspCysGlyAlaPheValLeuGl 296  
 Db 645 TGTACGAGAGTCTATTCCACAACAGAAAACACAGTACTGTGGAGTCTTTGTGCTCCA 704  
 QY 296 nTyrglyLysHisLeuAla-LeuSerGlnProPheSerPheThrGlnGlnAspMetPro 316  
 Db 705 GTACTGCAAGTCCCTCCGCTTAGAGCAGCCTTTCCAGTTTTCACAAGAAGACATGCC 764  
 QY 316 ysLeuArgArgGlnIleTyrglyLysGluLeuCysHisCysLysLeu 330  
 Db 765 GAGTCCGGAAGAGATTACAAAGAGCTATGTGAGTGCCTGCTC 808

RESULT 13

AAS25529/c

ID AAS25529 standard; cDNA; 503 BP.

XX AAS25529;

XX 07-NOV-2001 (first entry)

DE Human ovarian PCR-subtracted cDNA library clone #1614.

XX Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;

KW gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;  
 KW primer; probe.

OS Homo sapiens.

PN WO200157207-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US03733.

XX 04-FEB-2000; 2000US-0180403.

PR 28-MAR-2000; 2000US-0192745.

XX (CORI-) CORIXA CORP.

XX Algate PA, Mannion J;  
XX WPI; 2001-488879/53.  
XX New polynucleotides encoding ovarian tumour proteins, useful for  
PT treating ovarian cancer, and as probes, primers, and markers of cancer  
PT progression -  
XX Example 1; page 375; 378pp; English.  
XX The invention comprises compositions used for the therapy and diagnosis  
CC of ovarian cancer. The compositions comprise one or more ovarian tumour  
CC proteins, their associated polynucleotides, or immunogenic portions of  
CC the proteins. The ovarian tumour polynucleotides and polypeptides are  
CC useful for stimulating and/or expanding T cells specific for a tumour  
CC protein. They are also useful for inhibiting the development of cancer in  
CC a patient with an ovarian tumour DNA or protein by incubating isolated  
CC T-cells allowing them to proliferate, and administering to the patient.  
CC The sequences can be used as markers for cancer, for example, to monitor  
CC ovarian cancer progression. Probes and primers are useful in nucleic acid  
CC hybridisation, in detecting the presence of complementary sequences in a  
CC given sample, for preparing mutant species and for preparing other  
CC genetic constructions. Sequences AAS23820-AAS25231 and AAS25238-AAS25549  
CC represent human ovarian tumour protein cDNA clones.  
XX  
XX Sequence 503 BP; 120 A; 110 C; 147 G; 126 T; 0 other;

Alignment Scores:  
Pred. No.: 1.68e-67 Length: 503  
Score: 661.50 Matches: 125  
Percent Similarity: 88.03% Conservative: 0  
Best Local Similarity: 88.03% Mismatches: 0  
Query Match: 37.33% Indels: 17  
DB: 22 Gaps: 1

US-09-848-852A-3 (1-332) x AAS25529 (1-503)

QY 191 LeuArgThrLysGlyTyrAspGlyValLysArgTrpThrLysAsnValAspIlePheAsn 210  
DB 502 CTCGGTACCAAGGGTATGATGGGGTGAAGGTGGACCAAAACGGTGGACATCTTCAAT 443  
QY 211 LysGluLeuLeuLeuLeuProIleHisLeuGluValHisTrpSerLeuIleSerValAsp 230  
DB 442 AAGGAGCTACTGCTAATCCCATCCACCTGGAGGTGCTGTCCTCACTCTGTGAT 383  
QY 231 ValArgArgArgThrIleThrPheAspSerGlnArgThrLeuAsnArgCysPro 250  
DB 382 GTAGGCGAGCCACCATCACTATTTGACTCGCAGCGTACCCCTAAACCCGCGCTGCCCT 323  
QY 251 LysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHis 270  
DB 322 AAGCATATCCCAAGTATCTACAGGAGAGCGGGTAAAGAAAGACCGACTGGATTCCAC 263  
QY 271 GlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAsnAspSerAspCys 290  
DB 262 CAGGCTGGAAGGTACTTCAAATG----- 236  
QY 291 GlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerGlnPropheSerPheThr 310  
DB 235 -----TACTGCAAGCATCTGGCCCTGTCTCAGGCATTTCAGCTTCACC 194  
QY 311 GlnGlnAspMetProLysLeuArgArgGlnIleTyrLysGluLeuCysHisCysLysLeu 330  
DB 193 CACGAGACATGCCAAACTCTCGGCAGATCTACAAGGAGGTGTGTCACTGCAAACTC 134  
QY 331 ThrVal 332  
DB 133 ACTGTG 128  
RESULT 14  
ID AAI84371  
AAI84371 standard; cDNA; 358 BP.

XX AAI84371;  
XX AC  
XX DT 06-NOV-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 4431.  
XX DE  
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation; ss.  
XX OS Homo sapiens.  
XX PN WO200164835-A2.  
XX PD 07-SEP-2001.  
XX XX 26-FEB-2001; 2001WO-US04927.  
XX XX 28-FEB-2000; 2000US-0515126.  
XX PR 18-MAY-2000; 2000US-0577409.  
XX XX (HYSE-) HYSEQ INC.  
XX XX Tang YT, Liu C, Drmanac RT;  
XX WPI; 2001-514838/56.  
XX P-PSDB; AAO04440.  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX XX  
PS Claim 1; SEQ ID NO 4431; 1399pp + Sequence Listing; English.  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 358 BP; 93 A; 97 C; 89 G; 78 T; 1 other;  
Alignment Scores:  
Pred. No.: 1.04e-62 Length: 358  
Score: 618.50 Matches: 116  
Percent Similarity: 85.76% Conservative: 2  
Best Local Similarity: 85.29% Mismatches: 1  
Query Match: 34.90% Indels: 17  
DB: 22 Gaps: 1  
US-09-848-852A-3 (1-332) x AAI84371 (1-358)  
QY 195 GlyTyrAspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeu 214  
DB 2 GGTATGATGGGTGAAAAGGTGGACGACGAGACGGGACATCTTCAATAGGAGCTACTG 61  
QY 215 LeuIleProIleHisLeuGluValHisTrpSerLeuIleSerValAspValArgArg 234  
DB 62 CTAATCCCCATCCCTGGAGGTGCATTGCTCCTCATCTGTGTGATGAGGACCC 121  
QY 235 ThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgCysProLysHisIleAla 254  
DB 235 ThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgCysProLysHisIleAla 254

Db 122 ACCATCACCTATTTCGACTCGCAGGCTACCCCTAAACCGCGCTGCCCTAAGCATATTGCC 181  
 QY 255 LysTyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLys 274  
 Db 182 AGTATCTACAGCAGAGCGGTAAAGAACCCAGCTGGATTTCACCCAGGCGCTGGAAA 241  
 QY 275 GlyTyrPheLysMetAsnValAlaAlaArgGlnAsnAspSerAspCysGlyAlaPheVal 294  
 Db 242 GCTTACTTCAAAATG----- 256  
 QY 295 LeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMet 314  
 Db 257 -----TACTGCAACATCTGCCCTGTCTCAGCCATTTCAGCTTCACCCAGCAGACATG 310  
 QY 315 ProLysLeuArgGlnIleTyrLysGluLeuCysHisCysLysLeu 330  
 Db 311 CCCAACTCGTCGCAGATCTACAGGAGCTGTCTACTGCAAACTN 358  
 RESULT 15  
 AA193958/c  
 ID AA193958 standard; cDNA; 820 BP.  
 XX AC AA193958;  
 XX DT 13-NOV-2001 (first entry)  
 XX DE Human neuroblastoma expressed polynucleotide SEQ ID NO 33.  
 XX KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.  
 XX OS Homo sapiens.  
 XX PN WO200166719-A1.  
 XX PD 13-SEP-2001.  
 XX PF 02-MAR-2001; 2001WO-JP01629.  
 XX PR 07-MAR-2000; 2000JP-0159195.  
 XX PA (CHIB-) CHIBA PREFECTURE.  
 XX PI (HISM) HISAMITSU PHARM CO LTD.  
 XX PI Nakagawara A;  
 XX DR WPI; 2001-565584/63.  
 XX PT Nucleic acids originating in gene expressed in human neuroblastoma,  
 PT useful as probe or primer in diagnosing prognosis of human  
 PT neuroblastoma, malignancy and susceptibility indicator or tumour marker  
 PT for anti-cancer agents -  
 XX PS Claim 1; Page 59; 2979pp; Japanese.  
 XX CC The invention relates to novel genes (AA193926-AA197963) expressed in  
 CC human neuroblastoma. The nucleic acids are applicable as a probe or  
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and  
 CC susceptibility indicators or tumour markers for anti-cancer agents. The  
 CC gene information for diagnosing prognosis is related to factors similar  
 CC to that for N-myc and TrkA genes.  
 XX SQ Sequence 820 BP; 224 A; 223 C; 170 G; 191 T; 12 other;

Alignment Scores:  
 Pred. No.: 1.18e-61 Length: 820  
 Score: 614.00 Matches: 149  
 Percent Similarity: 72.07% Conservative: 11  
 Best Local Similarity: 67.12% Mismatches: 46  
 Query Match: 34.65% Indels: 19  
 DB: 22 Gaps: 5

US-09-848-852A-3 (1-332) x AA193958 (1-820)

QY 5 GlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHisLeuAlaProProAspAla 24  
 Db 784 CAACGCTTTTGGGGAA-ATTTGGCCCNAGGGGACGAGTTTGACCCCTTG---ATC 729  
 QY 25 SerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGlnGluLeuPheGln 44  
 Db 728 CAGATCCTTATCAGCAATNTTTCAGCATCGGGGACA--TNTGCCCAGGAGNT-TTTCAG 673  
 QY 45 GlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLysAlaGlyGlnHis 64  
 Db 672 GGTTCAGATT--GGCATGCAGAGGAGCAGAGCCCTGGGGAGAA---GCCGCCAGAA 617  
 QY 65 SerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAspGluPheLeuGln 84  
 Db 616 AGCCCTTTCNGAGAGGAGCATGTGATTCGCAACAGAGCATTTTGACGANTTCCTTCAA 557  
 QY 85 -ThrTyrGlySerLeuIleProLeuSerThrAspGluValGluLysLeuGlnAspIle 104  
 Db 556 AACGTATGGCAGCNTCATACCTTCAGCCTGATGAGGTAGTAGAAGTTGGAGGACAT 497  
 QY 104 ePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSe 124  
 Db 496 TTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGCTTGGTTCGACCTNATCCAGTC 437  
 QY 124 rTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHi 144  
 Db 436 TTACCAGCGGATGCCAGCAATGCCATGCTGAGGGGCTTCCGAGTGGCTTATAAGCGGCA 377  
 QY 144 sValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnVa 164  
 Db 376 CGTGCTGACCATGGATGACTTGGGACCTTGTATGGACAGAACTGGCTCAATGACCAGGT 317  
 QY 164 lMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVal----- 180  
 Db 316 GATGAACATGTATGGAGACCTGTGTCATGGACACACGTCCTGAAAGGTAGGCCCAACACAG 257  
 QY 181 -----HisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGln 195  
 Db 256 ATAGGTCAGTACCCAGAGGAACCTTCTGCAGTTT-----AAGCAGCTTTTAAAGCT 206  
 QY 195 yTyrAspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLe 215  
 Db 205 CCTTTTCATCTCTTTCATTTTACACAGAGGAGGTCTCTGTTTCAGGGAGAGAGGTGTAGT 146  
 QY 215 uIle 216  
 Db 145 GGA 142

Search completed: December 16, 2002, 21:27:46

Job time : 313 secs

**THIS PAGE BLANK (USPTO)**

Pred. No. is the number of results predicted by chance to have a

**JOURNAL**  
Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY





Db 1642 ATGTGCGCCAGGCGAATAATGACAGTGACTGGTGGCTTTTGTGTGCGTACTGCAAGC 1701  
QY 1440 ATCTGCGCCCTGTCTCAGCCATTTCAGCTTCACCCAGCAGGACATGCCAAACTTCGTGCGC 1499  
Db 1702 ATCTGCGCCCTGTCTCAGCCATTTCAGCTTCACCCAGCAGGACATGCCAAACTTCGTGCGC 1761  
QY 1500 AGATCTACAAGGAGCTGTCTCACTGCAAACTCACTGTGTGAGCCCTCGTACCCAGACCCC 1559  
Db 1762 AGATCTACAAGGAGCTGTCTCACTGCAAACTCACTGTGTGAGCCCTCGTACCCAGACCCC 1821  
QY 1560 AAGCCCATAAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAACTCCAGTTCTTTC 1619  
Db 1822 AAGCCCATAAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAACTCCAGTTCTTTC 1881  
QY 1620 CTCCTCTGGCTCTTCCCACTCACTTCCCTTGGTGTTCATATTTAAATGTTTCAATTTTC 1679  
Db 1882 CTCCTCTGGCTCTTCCCACTCACTTCCCTTGGTGTTCATATTTAAATGTTTCAATTTTC 1941  
QY 1680 TGTATTTTTCCTTTTGAGAGAATACCTTGTGATTTCTGATGTGCGAGGGGTGGCTACA 1739  
Db 1942 TGTATTTTTCCTTTTGAGAGAATACCTTGTGATTTCTGATGTGCGAGGGGTGGCTACA 2001  
QY 1740 GAAAAGCCCTTCTTCTCTGTTGCGAGGAGTGTGCGCCCTGTGCGCTGGTGAGCA 1799  
Db 2002 GAAAAGCCCTTCTTCTCTGTTGCGAGGAGTGTGCGCCCTGTGCGCTGGTGAGCA 2061  
QY 1800 GTCATCTCCCTTCCCGTGCAGGAGCAGGAATCAGTCTGGGGTGTGGCGGA 1859  
Db 2062 GTCATCTCCCTTCCCGTGCAGGAGCAGGAATCAGTCTGGGGTGTGGCGGA 2121  
QY 1860 CAATAGGATCACTGCTGCGAGATCTTCAAACTTTTATATATATATATATATATATATAT 1919  
Db 2122 CAATAGGATCACTGCTGCGAGATCTTCAAACTTTTATATATATATATATATATATATAT 2181  
QY 1920 AT 1979  
Db 2182 AT 2225  
QY 1980 TCAATAAAG 1988  
Db 2226 TCAATAAAG 2234

RESULT 2 AF199459 2258 bp mRNA linear PRI 31-OCT-2001  
LOCUS Homo sapiens SUMO-1 specific protease 3 (SSP3) mRNA, complete cds.  
DEFINITION AF199459  
ACCESSION AF199459.1 GI:16550942  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens.  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2258)  
AUTHORS Choi,S.J., Jeon,Y.J., Kim,K.I., Nishimori,S., Suzuki,T., Uchida,S., Shimbara,N., Tanaka,K. and Chung,C.H.  
DIRECT SUBMISSION  
TITLE Submitted (28-OCT-1999) Molecular Biology, Seoul National  
JOURNAL University, Shillim-dong san 56-1, Seoul 151-742, Korea  
FEATURES  
Location/Qualifiers  
source  
1..2258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..2258  
/gene="SSP3"  
71..1795  
/gene="SSP3"  
/codon\_start=1  
/product="SUMO-1 specific protease 3"  
1  
/protein\_id="AAL25652.1"  
/db\_xref="GI:16550943"  
/translation="MKETIQGTGSGPPEPGIPPAYSSPRRLRWPPPPKRLKS

GGGPGDPGSGTTPVARRLPVPRPSFDASAESEEEEEDEDEEEVAWRLPPRR  
SQTQSRRPSPRTHRTCSQRRRRAMRFLMYSKSTSLTTHWKLWGRGRRRR  
LAHPKHLSPQGGATPOVSPCCRFDPGRPPRLGLGLMAEDGVGRGPPVPSG  
PMEDGRLWTPKSPLDPSGLISCTLPNGFGGSGPEGRSLAPDASILLISNCSI  
GDVVAQELFQSGDLGMAEABRPEKAGQHSPLREHVTVCQSIIDFLQTYGSLIPL  
STDEVVEKLEDFQEFSPSRKGLVLQILOSYPMPGNVGRFVRYKHLVMTDD  
LGTLYGQNLNDQVNMNYDVLVMDTVPKVFHFFSFDYDKLRTKYGDGVKRWTKNVDI  
FNKELLIPIHLEVHWSLISDVMDTITFDNSFTLNRRCPKHLAKYLAQAAVKDR  
LDFHGWKGYFKMNVARQNDSDCGAFVLYQCKHLALSPQSFSTQDDMPKRLRQYKE  
LCHKLTV"  
BASE COUNT 519 a 626 c 618 g 495 t  
ORIGIN  
Query Match 96.7%; Score 1925.4; DB 9; Length 2258;  
Best Local Similarity 98.9%; Pred. No. 0;  
Matches 1967; Conservative 0; Mismatches 1; Indels 21; Gaps 2;  
QY 1 GCCTCCCTGTCCTCCCGACCTCTTTTGATGCCTAGCAAGTGAAGAGGAGGAGAGAGG 60  
Db 255 GCCTCCCTGTCCTCCCGACCTCTTTTGATGCCTAGCAAGTGAAGAGGAGGAGAGAGG 314  
QY 61 AGGAGGAGGAGATGAAGATGAAGAGGAGGAAGTGGCAGCTTGGAGGCTGCCCTCAAGAT 120  
Db 315 AGGAGGAGGAGATGAAGATGAAGAGGAGGAAGTGGCAGCTTGGAGGCTGCCCTCAAGAT 374  
QY 121 GGAGTCAGCTGGGAACCTCCCGAGGSCCGCCCTTCCCGCCCTCATCGAAAAACCT 180  
Db 375 GGAGTCAGCTGGGAACCTCCCGAGGSCCGCCCTTCCCGCCCTCATCGAAAAACCT 434  
QY 181 GCTCAGAGCCGCGCGCGAGCCATGAGAGCCCTCCGGATGCTGTCTACTCAAAAAGCA 240  
Db 435 GCTCAGAGCCGCGCGCGAGCCATGAGAGCCCTCCGGATGCTGTCTACTCAAAAAGCA 494  
QY 241 CCTCGCTGACATTCACACTGGAACTTTGGGGGGCCACCGGGGCGCGCGGGGCGCTCG 300  
Db 495 CCTCGCTGACATTCACACTGGAACTTTGGGGGGCCACCGGGGCGCGCGGGGCGCTCG 554  
QY 301 CACACCCCAAGAACCATCTTTCACCCCAAGAGGGGTGCGACGCCACAGGTGCCATCCC 360  
Db 555 CACACCCCAAGAACCATCTTTCACCCCAAGAGGGGTGCGACGCCACAGGTGCCATCCC 614  
QY 361 CTTGTGTGCTTTTGACTCCCCCGGGGCGCCACCTCCACCCCGGCTGTGCTAGGTG 420  
Db 615 CTTGTGTGCTTTTGACTCCCCCGGGGCGCCACCTCCACCCCGGCTGTGCTAGGTG 674  
QY 421 CTCCTATGCTCAGGATGGGTGAGAGGCTTCCACCACTGCTCGGGCCCCCATGG 480  
Db 675 CTCCTATGCTCAGGATGGGTGAGAGGCTTCCACCACTGCTCGGGCCCCCATGG 734  
QY 481 AGGAAGATGGACTCAGTGGACTCCAAAGTCTCTCTGACCCCTGACTCGGGCTCTTT 540  
Db 735 AGGAAGATGGACTCAGTGGACTCCAAAGTCTCTCTGACCCCTGACTCGGGCTCTTT 794  
QY 541 CATGTACTCTGCCAACCGTTTTTGGGGGACAATCTGGGCCAGAAAGGGAGCCCA-CTTGG 599  
Db 795 CATGTACTCTGCCAACCGTTTTTGGGGGACAATCTGGGCCAGAAAGGGAGCGAGCTTGG 854  
QY 600 CACCCCTCATGCCAGCATCCTCATCAGCAATGTGTGAGCATCGGGGACCATTGGCCCC 659  
Db 855 CACCCCTCATGCCAGCATCCTCATCAGCAATGTGTGAGCATCGGGGACCATTGGCCCC 914  
QY 660 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGGAGCGCTCGGGGAGA 719  
Db 915 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGGAGCGCTCGGGGAGA 974  
QY 720 AAGCCGGCCAGCACAGCCCCCTCGGAGAGGAGCATGTGACCTGCGGTACAGAGCATCTGG 779  
Db 975 AAGCCGGCCAGCACAGCCCCCTCGGAGAGGAGCATGTGACCTGCGGTACAGAGCATCTGG 1034  
QY 780 AGCAATTCCTCAAAGGTATGCGAGGCTCATACCCCTCAGCAGTGTAGGTAGTAGAGA 839  
Db 1035 AGCAATTCCTCAAAGGTATGCGAGGCTCATACCCCTCAGCAGTGTAGGTAGTAGAGA 1094

QY 840 AGCTGGAGGACATTTTCCAGCAGAGTTCACCCCTTCCAGGAGGCGCTGGTGTGC 899  
Db 1095 AGCTGGAGGACATTTTCCAGCAGAGTTCACCCCTTCCAGGAGGCGCTGGTGTGC 1154  
QY 900 AGCTGATCCAGTCTTACAGGGATGCCAGCAATGCCATGGTGAGGGCTTCGAGTGG 959  
Db 1155 AGCTGATCCAGTCTTACAGGGATGCCAGCAATGCCATGGTGAGGGCTTCGAGTGG 1214  
QY 960 CTTATAGCGCAGCTGCTGACCATGATGACTTGGGACCTGTATGGACAGAACTGCG 1019  
Db 1215 CTTATAGCGCAGCTGCTGACCATGATGACTTGGGACCTGTATGGACAGAACTGCG 1274  
QY 1020 TCAATGACCGAGGTGATGAACATGTATGGAGACCTGTGTCATGGACACAGTCCCTGAAAAGG 1079  
Db 1275 TCAATGACCGAGGTGATGAACATGTATGGAGACCTGTGTCATGGACACAGTCCCTGAAAAGG 1334  
QY 1080 TGCATTTCTCAATAGTTTCTTCTATGATAAATCCCTACAGGGTTATGATGGGGTGA 1139  
Db 1335 TGCATTTCTTCAATAGTTTCTTCTATGATAAATCCCTACAGGGTTATGATGGGGTGA 1394  
QY 1140 AAGGTGGACCAAAAGCGTGACATCTTCAATAGGAGCTACTGCTAATCCCATCCACC 1199  
Db 1395 AAGGTGGACCAAAAGCGTGACATCTTCAATAGGAGCTACTGCTAATCCCATCCACC 1454  
QY 1200 TGGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTAGGCGCAGCACCATCACTATTTCG 1259  
Db 1455 TGGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTAGGCGCAGCACCATCACTATTTCG 1514  
QY 1260 ACTCGAGCGTACCTTAACCGCGCTGCCCTTAAGCATATGCAAGTATCTACAGGCAG 1319  
Db 1515 ACTCGAGCGTACCTTAACCGCGCTGCCCTTAAGCATATGCAAGTATCTACAGGCAG 1574  
QY 1320 AGGCGGTAAAGAACCGACTGATTTCCACAGGCGTGAAGGTTACTTCAAAATGA 1379  
Db 1575 AGGCGGTAAAGAACCGACTGATTTCCACAGGCGTGAAGGTTACTTCAAAATGA 1634  
QY 1380 ATGTGGCCAGCGAGAATAATACAGTGACGTGTGTGCTTTTGTGTTGCAGTACTGCAAGC 1439  
Db 1635 ATGTGGCCAGCGAGAATAATACAGTGACGTGTGTGCTTTTGTGTTGCAGTACTGCAAGC 1694  
QY 1440 ATCTGGCCCTGTCTACGCCATTCAGCTTACCAGCAGGACATGCCCAAACTTCGTGCGG 1499  
Db 1695 ATCTGGCCCTGTCTACGCCATTCAGCTTACCAGCAGGACATGCCCAAACTTCGTGCGG 1754  
QY 1500 AGATCTACAAGGAGCTGTGCTACTGCAAACTCACTGTGTCAGCTCTGCTACCCAGACCCC 1559  
Db 1755 AGATCTACAAGGAGCTGTGCTACTGCAAACTCACTGTGTCAGCTCTGCTACCCAGACCCC 1814  
QY 1560 AAGCCCAATAATGGAGGAGACATGGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC 1619  
Db 1815 AAGCCCAATAATGGAGGAGACATGGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC 1874  
QY 1620 CTCTCTGCTCTTCCGACATCTCCCTTGGTTTTCATATTTAAATGTTTCAATTC 1679  
Db 1875 CTCTCTGCTCTTCCGACATCTCCCTTGGTTTTCATATTTAAATGTTTCAATTC 1934  
QY 1680 TGTATTTTCTTTTGTGAGAGAATACTCTGATTTCTCATGTCGACGGGGTGGCTACA 1739  
Db 1935 TGTATTTTCTTTTGTGAGAGAATACTCTGATTTCTCATGTCGACGGGGTGGCTACA 1994  
QY 1740 GAAAGCCCTTCTTCTCTCTGTTGAGGGAGTGTGGCCCTGTGCGCTGGGTGGAGCA 1799  
Db 1995 GAAAGCCCTTCTTCTCTCTGTTGAGGGAGTGTGGCCCTGTGCGCTGGGTGGAGCA 2054  
QY 1800 GTCATCTCCCTTCCCTGTGAGGAGCAGGAAATCACTGCTGGGGTGGTGGCGGA 1859  
Db 2055 GTCATCTCCCTTCCCTGTGAGGAGCAGGAAATCACTGCTGGGGTGGTGGCGGA 2114  
QY 1860 CAATAGATCACTGCTGCCAGATCTTCAACCTTTTATATATATATATATATATATAT 1919  
Db 2115 CAATAGATCACTGCTGCCAGATCTTCAACCTTTTATATATATATATATATATATATAT 2154  
QY 1920 AT 1979

Db 2155 AT 2214  
QY 1980 TCAATAAG 1988  
Db 2215 TCAATAAG 2223  
RESULT 3  
AY008763 2224 bp mRNA linear PRI 21-NOV-2000  
LOCUS Homo sapiens sentrin/SUMO-specific protease (SEN3) mRNA, complete.  
DEFINITION cds.  
ACCESSION AY008763  
VERSION AY008763.1 GI:11245810  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2224)  
AUTHORS Yeh,E.T., Gong,L. and Kamitani,T.  
TITLE Ubiquitin-like proteins: new wines in new bottles  
JOURNAL Gene 248 (1-2), 1-14 (2000)  
MEDLINE 20267842  
PUBMED 10806345  
REFERENCE 2 (bases 1 to 2224)  
AUTHORS Gong,L. and Yeh,E.T.  
TITLE Direct Submission  
JOURNAL Submitted (04-OCT-2000) Institute of Molecular Medicine, University of Texas-Houston, 2121 W. Holcombe Blvd. 928, Houston, TX 77030, USA  
FEATURES  
source Location/Qualifiers  
1..2224  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17p13"  
1..2224  
/gene="SEN3"  
52..1776  
/gene="SEN3"  
/codon\_start=1  
/product="sentrin/SUMO-specific protease"  
/protein\_id="AAG33252.1"  
/db\_xref="GI:11245811"  
/translation="MKETIQGTSNGPEPPGIPPAYSSPRRDGLRWPPPKPLKLS  
GGFGPDGSGTTPTRRLPAPRPSFDASASEEEEEDEEEDEEEVAARLPRW  
SOLGTSQRPALRPSHRKTSQRRRRAMRAFQMLLYSKSTSLTHFKLWGRHRRRG  
LAHPKNHLSPOCGATPOVSPCCRFDSPPRPPRLGLLGMALMADVGRGPPVPSG  
PMWEDGLRWTPKSPLDPDGSLLSCTLPNGFGSGSGPEGERSLAPDASILISNVC  
GDHVAQELFGSDLGMAEAEAPRPEKAGHSPLEHVTVCQSILDEFLTQYGLIPL  
STDEVVEKLEIFQEFSTPSRRGLVLQISQIRMPGNAMVGRVATKRHLVTMD  
LGTLYQNMNDQVNMGLVMDVTPKVFHFNFFYDKLRGTGDKYKRWTKNVDI  
FNKEILLPIHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIKYLQAEAVKDR  
LDFHOGWKGYFKMNVARQNDSDCGAFVLQYCKHLALSQPFSTOODMPKLRQIYE  
LCHCKLTV"  
BASE COUNT 502 a 615 c 612 g 495 t  
ORIGIN  
Query Match 95.8%; Score 1907.8; DB 9; Length 2224;  
Best Local Similarity 98.3%; Pred. No. 0;  
Matches 1956; Conservative 0; Mismatches 12; Indels 21; Gaps 2;  
QY 1 GCCTCCTGTCCCGGACCCCTCTTTTGTAGTCCTCAGACAGTGAAGGAGGAGGAAGAGG 60  
Db 236 GCCTCCTGTCCCGGACCCCTCTTTTGTAGTCCTCAGACAGTGAAGGAGGAGGAAGAGG 295  
QY 61 AGCAGGAGGAGGATGAAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120  
Db 296 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 355  
QY 121 GGAGTCAGCTGGGAACCTCCCGAGCGCGCCCTTCCCGCGCCCACTCATCGAAAAACCT 180

|||||  
Db 356 GGAGTCAGTGGGAACCTCCACAGCGCCCGCGCTCTCCGACCCCTCATAGAAAACCT 415  
QY 181 GCTCAGCGCGCGCGAGCCATCAGAGCCTTCGGATGTCTCTACTCAAAAAGCA 240  
Db 416 GCTCAGCGCGCGCGAGCCATCAGAGCCTTCGAGTGTCTACTCAAAAAGCA 475  
QY 241 CCTCGCTGACATTCACATCGGAAGCTTTGGGGGCGCCACCGGGGCGGGCGGGGCGCTCG 300  
Db 476 CCTCGCTGACATTCACATCGGAAGCTTTGGGGGCGCCACCGGGGCGGGCGGGGCGCTCG 535  
QY 301 CACACCCCAAGAACATCTTCCACCCCAAGAGGGGTGCGACGCCACAGGTGCGCATCC 360  
Db 536 CACACCCCAAGAACATCTTCCACCCCAAGAGGGGTGCGACGCCACAGGTGCGCATCC 595  
QY 361 CCTGTTGTGCTTTGACTCCCGCGGGGCGCCACCTCCACCGCGGCTGGGTCTGCTAGGTG 420  
Db 596 CCTGTTGTGCTTTGACTCCCGCGGGGCGCCACCTCCACCGCGGCTGGGTCTGCTAGGTG 655  
QY 421 CTCTCATGGCTGAGATGGGGTGAGAGGTCTTCCACCAAGTCCCTCTGGGGCGCCCATCG 480  
Db 656 CTCTCATGGCTGAGATGGGGTGAGAGGTCTTCCACCAAGTCCCTCTGGGGCGCCCATCG 715  
QY 481 AGGAAGATGGACTAGGTGGACTCCAAAGTCTCTCTTGGAACCTGACTCGGGGCTCTCTTT 540  
Db 716 AGGAAGATGGACTAGGTGGACTCCAAAGTCTCTCTTGGAACCTGACTCGGGGCTCTCTTT 775  
QY 541 CATGTACTCTGCCCAACGTTTTGGGGGACATCTGGGCCAGAGGGGAGCCCA-CTTGG 599  
Db 776 CATGTACTCTGCCCAACGTTTTGGGGGACATCTGGGCCAGAGGGGAGCGACTTGG 835  
QY 600 CACCCCTCTGATGCCAGCATCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGGCC 659  
Db 836 CACCCCTCTGATGCCAGCATCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGGCC 895  
QY 660 AGGAGCTTTTCAGGGCTCAGATTTGGGATGGGATGGGAGAGAGGAGAGGCGCTGGGGAGA 719  
Db 896 AGGAGCTTTTCAGGGCTCAGATTTGGGATGGGATGGGAGAGAGGAGAGGCGCTGGGGAGA 955  
QY 720 AAGCGGGGAGCAGACAGCCCTCGCAGAGGAGCATGTGACCTCGCTACAGAGCATCTTGG 779  
Db 956 AAGCGGGGAGCAGACAGCCCTCGCAGAGGAGCATGTGACCTCGCTACAGAGCATCTTGG 1015  
QY 780 AGGAATTCCTTCAACGATGCGACGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA 839  
Db 1016 AGGAATTCCTTCAACGATGCGACGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA 1075  
QY 840 AGCTGAGGACATTTCCAGCAGGAGTTTCCACCCCTTCCAGGAGGCGCTGGTGTTC 899  
Db 1076 AGCTGAGGACATTTCCAGCAGGAGTTTCCACCCCTTCCAGGAGGCGCTGGTGTTC 1135  
QY 900 AGCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGTGAGGGGCTTCCGAGTGG 959  
Db 1136 AGCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGTGAGGGGCTTCCGAGTGG 1195  
QY 960 CTATAAGCGGACGCTGTGACCATGGATGATGCTGGGACCTTGTATGACAGAACTGGC 1019  
Db 1196 CTATAAGCGGACGCTGTGACCATGGATGATGCTGGGACCTTGTATGACAGAACTGGC 1255  
QY 1020 TCAATGACAGGTGATGAACATGTATGAGACCTGGTATGGACACAGTCCCTGAAAAGG 1079  
Db 1256 TCAATGACAGGTGATGAACATGTATGAGACCTGGTATGGACACAGTCCCTGAAAAGG 1315  
QY 1080 TGCATTTCTTCAATAGTTTCTTATGATAAACTCCGTACCAAGGTTATGATGGGTGA 1139  
Db 1316 TGCATTTCTTCAATAGTTTCTTATGATAAACTCCGTACCAAGGTTATGATGGGTGA 1375  
QY 1140 AAAGTGGACCAAAAACGTGGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACC 1199  
Db 1376 AAAGTGGACCAAAAACGTGGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACC 1435  
QY 1200 TGGAGGTGATTTGGTCCCTCATCTCTGTGTGATGTGAGGGGAGCGACCATCATCTTTTG 1259  
|||||

Db 1436 TGGAGTGTCATTTGGTCCCTCATCTCTGTTGATGTGAGGCGAGCCACCATCATCTTTTG 1495  
QY 1260 ACTCGCAGGTACCTTAAACCGCGCTGCCCTAAGCATATTCGCAAGTATCTACAGCGAG 1319  
Db 1496 ACTCGCAGGTACCTTAAACCGCGCTGCCCTAAGCATATTCGCAAGTATCTACAGCGAG 1555  
QY 1320 AGCGGTAAAGAAAGACCGGACTTTCACAGGCGTGGAAAGGTTACTTCAAATGA 1379  
Db 1556 AGCGGTAAAGAAAGACCGGACTTTCACAGGCGTGGAAAGGTTACTTCAAATGA 1615  
QY 1380 ATGTGCCAGGAGCAATATGACAGTACTGTGGTCTTTTGTGTGAGTACTGCAAGC 1439  
Db 1616 ATGTGCCAGGAGCAATATGACAGTACTGTGGTCTTTTGTGTGAGTACTGCAAGC 1675  
QY 1440 ATCTGGCCGTCTCAGCCATTCAGCTTCACCCAGCAGACATGCCAAACTTCGTCGCG 1499  
Db 1676 ATCTGGCCGTCTCAGCCATTCAGCTTCACCCAGCAGACATGCCAAACTTCGTCGCG 1735  
QY 1500 AGATCTACAAGAGCTGTGTCACTGCAAACTCAGTGTGTGAGCCTCGTACCCAGACCCC 1559  
Db 1736 AGATCTACAAGAGCTGTGTCACTGCAAACTCAGTGTGTGAGCCTCGTACCCAGACCCC 1795  
QY 1560 AAGCCCAATAATGGGAAGGAGACATGGAGTCCCTTCCCAAGAAACTCCAGTTCCTTTC 1619  
Db 1796 AAGCCCAATAATGGGAAGGAGACATGGAGTCCCTTCCCAAGAAACTCCAGTTCCTTTC 1855  
QY 1620 CTCTCTGGCTCTCCCACTCACTTCCCTTGTGTTTTCATATTTAAATGTTTCAATTC 1679  
Db 1856 CTCTCTGGCTCTCCCACTCACTTCCCTTGTGTTTTCATATTTAAATGTTTCAATTC 1915  
QY 1680 TGTATTTTCTTTTGAGAGAATCTTGTGATTTCTGATGTGAGGGGTGGCTACA 1739  
Db 1916 TGTATTTTCTTTTGAGAGAATCTTGTGATTTCTGATGTGAGGGGTGGCTACA 1975  
QY 1740 GAAAAGCCCTTCTTCTCTCTTTGAGGGGAGTGTGGCCCTGTGGCTGGGAGCA 1799  
Db 1976 GAAAAGCCCTTCTTCTCTCTTTGAGGGGAGTGTGGCCCTGTGGCTGGGAGCA 2035  
QY 1800 GTCATCTCCCTTCCCGTGCGAGGAGCAGAAATCAGTGTGGGGGTGGTGGCGGA 1859  
Db 2036 GTCATCTCCCTTCCCGTGCGAGGAGCAGAAATCAGTGTGGGGGTGGTGGCGGA 2095  
QY 1860 CAATAGGATCACTGCTGCCAGATCTTCAAACCTTTTATATATATATATATATATAT 1919  
Db 2096 CAATAGGATCACTGCTGCCAGATCTTCAAACCTT-----TATAT 2135  
QY 1920 AT 1979  
Db 2136 AT 2195  
QY 1980 TCAATAAAG 1988  
Db 2196 TCAATAAAG 2204

RESULT 4  
AX081143  
LOCUS AX081143 2206 bp DNA linear PAT 27-FEB-2001  
DEFINITION Sequence 9 from Patent WO0109292.  
ACCESSION AX081143  
VERSION AX081143.1 GI:13170039  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Gong,L. and Yeh,E.T.  
TITLE Composition and methods relating to senpi - a sentrin-specific  
protease  
JOURNAL Patent: WO 0109292-A 9 08-FEB-2001;  
Board of Regents, The University of Texas System (US)  
FEATURES  
Location/Qualifiers

source	1. .2206		/organism="Homo sapiens"		/db_xref="taxon:9606"			
BASE COUNT	491 a	612 c	605 g	498 t				
ORIGIN								
Query Match	92.6%;		Score	1843.4;	DB	6;	Length	2206;
Best Local Similarity	96.8%;		Pred. No.	0;				
Matches 1925;	Conservative	0;	Mismatches	31;	Indels	33;	Gaps	
QY	1	GCCTCCTGTC	CCCGACCTCTTTT	GATGCTCAGCAAGTGAAGAGGAGGAAGAGG	60			
Db	230	GCCTCCTGTC	CCCGCCCATCTTTT	GATGCTCAGCTAGTGAAGAAGAGGAGGAAG	289			
QY	61	AGGAGGAGGAG	GATGAAGATGAAGAGGAGGAAGTGGCAGCTTGGAGGCTGCCCCAAGAT	120				
Db	290	AGGAG-----	GAAGATGAGGAGGAAGTAGCAGCTTGGAGGCTACCCCTAGGT	337				
QY	121	GGAGTCAGCT	GGGAACCTCCCGAGCGCCCGCCCTTCCCGCCCCACTCATCGAAAAACCT	180				
Db	338	GGGCGCAACT	GGGGCCTCCCGAGGCTCTCGAGCTCTCCGACCTCTCATAGAAAAACCT	397				
QY	181	GCTCAGAGCG	CGCGCGGAGGCATGAGAGCCTTCGGGATGCTGCTCTACTCTAAAAAGCA	240				
Db	398	GCTCAGAGCG	CGCGCGGAGGCATGAGAGCCTTCGATGCTGCTCTACTCTAAAAAGCA	457				
QY	241	CCTCGCTGAC	ATTCCACTGGAAGCTTTGGGGGCGCCACCGGGCGCGCGGGCGCTCG	300				
Db	458	CCTCGCTGAC	ATTCCACTGGAAGCTTTGGGGGCGCCACCGGGCGCGCGGGCGCTCG	517				
QY	301	CACACCCCAAG	AACCATCTTTTCAACCCAGCAAGGGGTGGACGCCACAGGTGCCATCCC	360				
Db	518	CACACCCCAAG	AACCATCTTTTCAACCCAGCAAGGGGTGGACGCCACAGGTGCCATCCC	577				
QY	361	CCTGTTGTCG	TTTTGACTCCCGCGGGGCGCCACCTCCACCCCGGCTGGGTCTGCTAGGTG	420				
Db	578	CCTGTTGTCG	TTTTGACTCCCGCGGGGCGCCACCTCCACCCCGGCTGGGTCTGCTAGGTG	637				
QY	421	CTCTCATGCT	GAGGATGGGTGAGAGGTCTCCACAGTGCCTCTGGGCCCGCCCATGG	480				
Db	638	CTCTCATGCT	GAGGATGGGTGAGAGGTCTCCACAGTGCCTCTGGGCCCGCCCATGG	697				
QY	481	AGGAAGATG	GACTCAGGTGGACTCCAAAGTCTCCTCTGGACCCCTGACTCGGGCCTCCTTT	540				
Db	698	AGGAAGATG	GACTCAGGTGGACTCCAAAGTCTCCTCTGGACCCCTGACTCGGGCCTCCTTT	757				
QY	541	CATGTACTCT	GCCCAAGGTTTTGGGGGACAATCTGGGCCAGAGGGAGCGCA-CTTGG	599				
Db	758	CATGTACTCT	GCCCAAGGTTTTGGGGGACAATCTGGGCCAGAGGGAGCGCAGCTTGG	817				
QY	600	CACCCCTGAT	GCCAGCATCCTCATCAGCAATGTCTGAGCATCGGGAGCATGTGGCCC	659				
Db	818	CACCCCTGAT	GCCAGCATCCTCATCAGCAATGTCTGAGCATCGGGAGCATGTGGCCC	877				
QY	660	AGGAGCTTTT	TTCAGGGCTCAGATTTGGGCATGGCAGAGGCGCAGAGAGCCTGGGGAGA	719				
Db	878	AGGAGCTTTT	TTCAGGGCTCAGATTTGGGCATGGCAGAGGCGCAGAGAGCCTGGGGAGA	937				
QY	720	AAGCCGGCAG	CACAGCCCGCTGGGAGAGGAGCATGTGACCTCGGTACAGAGCATCTTGG	779				
Db	938	AAGCCGGCAG	CACAGCCCGCTGGGAGAGGAGCATGTGACCTCGGTACAGAGCATCTTGG	997				
QY	780	ACGAATTCCT	TCAACAGTATGGCAGCTCATACCCCTCAGCATGTAGAGTAGTAGAGA	839				
Db	998	ACGAATTCCT	TCAACAGTATGGCAGCTCATACCCCTCAGCATGTAGAGTAGTAGAGA	1057				
QY	840	AGCTGGAGG	ACATTTTCCAGCAGGAGTTTTCACCCCTTCCAGGAAGGGCCTGGTGTTC	899				
Db	1058	AGCTGGAGG	ACATTTTCCAGCAGGAGTTTTCACCCCTTCCAGGAAGGGCCTGGTGTTC	1117				
QY	900	AGCTGATCC	AGTCTTACACAGCGGATGCCAGGCAATGCCATGGTGGAGGGCTCCGAGTGG	959				
Db	1118	AGCTGATCC	AGTCTTACACAGCGGATGCCAGGCAATGCCATGGTGGAGGGCTCCGAGTGG	1177				

QY	960	CTTATAGCGGCACGTGCTGACCATGGATGACCTTGGGACCTTGTATGGACAGAACTGGC	1019						
DB	1178	CTTATAGCGGCACGTGCTGACCATGGATGACCTTGGGACCTTGTATGGACAGAACTGGC	1237						
QY	1020	TCATAGCAGGCTGATGAACATGTATGGACACCTGTGTCATGGACACAGTCCCTGAAAAGG	1079						
DB	1238	TCATAGCAGGCTGATGAACATGTATGGACACCTGTGTCATGGACACAGTCCCTGAAAAGG	1297						
QY	1080	TGCATTTCTTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGGTGA	1139						
DB	1298	TGCATTTCTTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGGTGA	1357						
QY	1140	AAAGGTGGACCAAAAACGTGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACC	1199						
DB	1358	AAAGGTGGACCAAAAACGTGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACC	1417						
QY	1200	TGGAGTGGCATTTGGTCCCTCATCTCTGTTGATGTAGGGCAGCACCACATCACTATTTTG	1259						
DB	1418	TGGAGTGGCATTTGGTCCCTCATCTCTGTTGATGTAGGGCAGCACCACATCACTATTTTG	1477						
QY	1260	ACTCGAGCGTACCTAAACCGCGCTGCCCCTAAGCATATTGCCAAGTATCTACAGCGAG	1319						
DB	1478	ACTCGAGCGTACCTAAACCGCGCTGCCCCTAAGCATATTGCCAAGTATCTACAGCGAG	1537						
QY	1320	AGGCGGTAAAGAAAGACCGACTGGATTTCACCCAGGGCTGGAAGGTTACTTCAAAATGA	1379						
DB	1538	AGGCGGTAAAGAAAGACCGACTGGATTTCACCCAGGGCTGGAAGGTTACTTCAAAATGA	1597						
QY	1380	ATGTGCCAGGCGAGAAATTAATGACAGTGACTGTGTGCTTTGTTGTGTCAGTACTGCAAGC	1439						
DB	1598	ATGTGCCAGGCGAGAAATTAATGACAGTGACTGTGTGCTTTGTTGTGTCAGTACTGCAAGC	1657						
QY	1440	ATGTGCCCTGTCTCAGCCATTGAGCTTACCAGCAGGAGACATGCCCAAACTTCGTCGGC	1499						
DB	1658	ATGTGCCCTGTCTCAGCCATTGAGCTTACCAGCAGGAGACATGCCCAAACTTCGTCGGC	1717						
QY	1500	AGATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTGTGAGCCTCGTACCCAGAGCCCC	1559						
DB	1718	AGATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTGTGAGCCTCGTACCCAGAGCCCC	1777						
QY	1560	AAGCCATAAATGGGAGGAGACATGGGAGTCCCTTCCCAAGAACTCAGTTCCTTTC	1619						
DB	1778	AAGCCATAAATGGGAGGAGACATGGGAGTCCCTTCCCAAGAACTCAGTTCCTTTC	1837						
QY	1620	CTCTCTGCTCTTCCCACTCACTTCCCTTGGTTTTCATATTAAATCTTCAATTTTC	1679						
DB	1838	CTCTCTGCTCTTCCCACTCACTTCCCTTGGTTTTCATATTAAATCTTCAATTTTC	1897						
QY	1680	TGTATTTTCTTTTCTTTTGAGAGAAATCTTGTGATTCTGATGTGAGGGGGTGGCTACA	1739						
DB	1898	TGTATTTTCTTTTCTTTTGAGAGAAATCTTGTGATTCTGATGTGAGGGGGTGGCTACA	1957						
QY	1740	GAAGAGCCCTTCTTCTCTCTGTTTGCAGGGAGTGTGGCCCTGTGGCTGGGGTGGAGCA	1799						
DB	1958	GAAGAGCCCTTCTTCTCTCTGTTTGCAGGGAGTGTGGCCCTGTGGCTGGGGTGGAGCA	2017						
QY	1800	GTCACTCTCCCTCTCCCGTGCAGGAGCAGGAATCAGTCTGGGGTGGTGGGGGA	1859						
DB	2018	GTCACTCTCCCTCTCCCGTGCAGGAGCAGGAATCAGTCTGGGGTGGTGGGGGA	2077						
QY	1860	CAATAGGATCACTGCTGCCAGATCTTCAAACTTTTATATATATATATATATATATAT	1919						
DB	2078	CAATAGGATCACTGCTGCCAGATCTTCAAACTTTTATATATATATATATATATATAT	2117						
QY	1920	AT	1979						
DB	2118	AT	2177						
QY	1980	TCAATAAAG	1988						
DB	2178	TCAATAAAG	2186						

RESULT 5	HSMB00198	1782 bp	mRNA	linear	PRI 10-MAR-2001
LOCUS	HSMB00198	1782 bp	mRNA	linear	PRI 10-MAR-2001
DEFINITION	Homo sapiens mRNA; cDNA DKFZp586K0919 (from clone DKFZp586K0919); complete cds.				
ACCESSION	AL050283				
VERSION	AL050283.1	GI:4886466			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Wiemann, S., Weil, B., Wellenreuther, R., Gassenhuber, J., Glassl, S., Ansong, W., Boecker, M., Bloeker, H., Bauersachs, S., Blum, H., Lauber, J., Duesterhoeft, A., Beyer, A., Koehler, K., Strack, N., Mewes, H. W., Othenwaelder, B., Obermaier, B., Tampe, J., Heubner, D., Wambutt, R., Korn, B., Klein, M. and Poustka, A.				
TITLE	Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs				
JOURNAL	Genome Res. 11 (3), 422-435 (2001)				
PUBMED	11230166				
REFERENCE	2 (bases 1 to 1782)				
AUTHORS	Wambutt, R., Heubner, D., Mewes, H. W., Gassenhuber, J. and Wiemann, S.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-MAR-1999) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY				
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp586K0919) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/.				
FEATURES	Location/Qualifiers				
source	1. .1782				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="DKFZp586K0919"				
	/tissue_type="uterus"				
	/clone_lib="586 (synonym: hutel). Vector pSport1; host DH10B; sites NotI + SalI/MluI"				
	/dev_stage="adult"				
gene	204..1319				
	/gene="DKFZp586K0919"				
CDS	204..1319				
	/gene="DKFZp586K0919"				
	/note="similarity to S.pombe SPBC19G7.09"				
	/codon_start=1				
	/product="hypothetical protein"				
	/protein_id="CAB43384.1"				
	/db_xref="GI:4886467"				
	/db_xref="SPTREMBL:Q9Y3W9"				
	/translation="NAEDVGRGSPVPSPMEEDGLRWTPKSPDPDPSGLLSCTLPN GFGGSGPGEERSLAPDASILLISNVCISGDHVAQELFGGSDLSNAERPERGKAGQ HSPLEEHVTCQVSLDEFLTQYGLIFLSTDEVVEKLEDFQQEFSPSRGLVLQL IQSYRPMGNAMVRGFRVAYKRHLVTMDDLGLTGQNLNDQVMNMGDLYNDVTYPEK VHFNSFFYDLKRTGYDGVKRVKTNVDIFNKLILLIPILHEVHWSLISVDYRRRTIT YFDQSRTNRRCPHIAKYALQAEAVKDRDLDFHOGWKGYFKMNVARQNNDSDCGAFVL QYCKHLALSQPSFTQDDMPKLRRQIYKELCHCKLTV"				
polyA_signal	1741..1746				
polyA_site	1745				
BASE COUNT	419 a	466 c	475 g	422 t	
ORIGIN					
Query Match	85.0%; Score 1692.4; DB 9; Length 1782;				
Best Local Similarity	98.7%; Pred. No. 0;				
Matches 1734; Conservative	0; Mismatches 1; Indels 21; Gaps 2;				
QY	234	AAAAGCACCTCGCTGACATTCACACTGGAAGCTTTGGGGCGCCACCGGGCGCGCGG	CGG	CGG	CGG

Db	12	AAAAGCACCTCGCTGACATTCACACTGGAAGCTTTGGGGCGCCACCGGGCGCGCGG	CGG	CGG	CGG
QY	294	GGCCTCGCACACCCCAAGAACCATCTTTTACCCCAAGAGGGGTGCGAGCCACAGGTG	CGG	CGG	CGG
Db	72	GGCCTCGCACACCCCAAGAACCATCTTTTACCCCAAGAGGGGTGCGAGCCACAGGTG	CGG	CGG	CGG
QY	354	CCATCCCTCGCTGCTGCTTTGACTCCCTCCCGGGGGCCACCTCCACCCCGGCTGGTCTG	CGG	CGG	CGG
Db	132	CCATCCCTCGCTGCTGCTTTGACTCCCTCCCGGGGGCCACCTCCACCCCGGCTGGTCTG	CGG	CGG	CGG
QY	414	CTAGTGTCTCTCATGGCTGAGGATGGGTGAGAGGTCTTCCACCACTGCTTGGGGCCC	CGG	CGG	CGG
Db	192	CTAGTGTCTCTCATGGCTGAGGATGGGTGAGAGGTCTTCCACCACTGCTTGGGGCCC	CGG	CGG	CGG
QY	474	CCCATGGAGGAAGATGAGATCAGGTGAGTCCAAAGTCTCTCTGACCTGACTCGGGC	CGG	CGG	CGG
Db	252	CCCATGGAGGAAGATGAGATCAGGTGAGTCCAAAGTCTCTCTGACCTGACTCGGGC	CGG	CGG	CGG
QY	534	CTCCTTTCTATCTCTGCCCAACGGTTTGGGGGACAACTCTGGGCGCAGAAGGGAGCGC	CGG	CGG	CGG
Db	312	CTCCTTTCTATCTCTGCCCAACGGTTTGGGGGACAACTCTGGGCGCAGAAGGGAGCGC	CGG	CGG	CGG
QY	594	A-CTTGGCACCCCTGATGCCAGATCTCTATCAGCAATGTGTGACGATTCGGGGACCAT	CGG	CGG	CGG
Db	372	AGCTTGGCACCCCTGATGCCAGATCTCTATCAGCAATGTGTGACGATTCGGGGACCAT	CGG	CGG	CGG
QY	653	GTGGCCCGAGAGCTTTTTCAGGGCTCAGATTGGGGATGGCAGAGAGAGAGCCCT	CGG	CGG	CGG
Db	432	GTGGCCCGAGAGCTTTTTCAGGGCTCAGATTGGGGATGGCAGAGAGAGAGCCCT	CGG	CGG	CGG
QY	713	GGGGAGAAAGCGCGCCAGCACAGCCCTGCGAGAGAGGAGCATGTGACCTGCGTACAGAGC	CGG	CGG	CGG
Db	492	GGGGAGAAAGCGCGCCAGCACAGCCCTGCGAGAGAGGAGCATGTGACCTGCGTACAGAGC	CGG	CGG	CGG
QY	773	ATCTTGGACGAATTCCTTCAAAGTATGGCAGCTCATACCCCTCAGCACTGTATGAGGTA	CGG	CGG	CGG
Db	552	ATCTTGGACGAATTCCTTCAAAGTATGGCAGCTCATACCCCTCAGCACTGTATGAGGTA	CGG	CGG	CGG
QY	833	GTAGAGAACTGGAGGACATTTTCCAGCAGGAGTTTTCACCCCTTCCAGGAGGGCCCTG	CGG	CGG	CGG
Db	612	GTAGAGAACTGGAGGACATTTTCCAGCAGGAGTTTTCACCCCTTCCAGGAGGGCCCTG	CGG	CGG	CGG
QY	893	GTGTTGCAGCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGTGAGGGCTTC	CGG	CGG	CGG
Db	672	GTGTTGCAGCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGTGAGGGCTTC	CGG	CGG	CGG
QY	953	CGAGTGGCTTATAAGCGGCACGCTGTGACCATGGATGACTTGGGGACCTTGTATGGACAG	CGG	CGG	CGG
Db	732	CGAGTGGCTTATAAGCGGCACGCTGTGACCATGGATGACTTGGGGACCTTGTATGGACAG	CGG	CGG	CGG
QY	1013	AACTGGCTCAATGACCCAGGTGATGAACATGTATGGAGACCTGGTATGACACAGTCCCT	CGG	CGG	CGG
Db	792	AACTGGCTCAATGACCCAGGTGATGAACATGTATGGAGACCTGGTATGACACAGTCCCT	CGG	CGG	CGG
QY	1073	GAAGAGTGCATTTCTTCAATAGTTTCTTATGATAAACTCCGTACCAGGGTTATGAT	CGG	CGG	CGG
Db	852	GAAGAGTGCATTTCTTCAATAGTTTCTTATGATAAACTCCGTACCAGGGTTATGAT	CGG	CGG	CGG
QY	1133	GGGGTGAAGAGTGGACCAAAAACGTGGACATCTTCAATAAGAGAGTACTGTCTAAATCCC	CGG	CGG	CGG
Db	912	GGGGTGAAGAGTGGACCAAAAACGTGGACATCTTCAATAAGAGAGTACTGTCTAAATCCC	CGG	CGG	CGG
QY	1193	ATCCACCTGGAGTGCATTTGGTGGCTTCATCTCTGTGTATGTGAGGGAGCGCACCATCAC	CGG	CGG	CGG
Db	972	ATCCACCTGGAGTGCATTTGGTGGCTTCATCTCTGTGTATGTGAGGGAGCGCACCATCAC	CGG	CGG	CGG
QY	1253	TATTTTGACTCGCAGGTACCTTAAACCGCGCTGCTCCCTAAAGCATATTGCCAAGTATCTA	CGG	CGG	CGG
Db	1032	TATTTTGACTCGCAGGTACCTTAAACCGCGCTGCTCCCTAAAGCATATTGCCAAGTATCTA	CGG	CGG	CGG
QY	1313	CAGGCAGAGCGGTAAAGAAAGAACCCGACTGGATTTCACACAGGCTGGAAGGTTACTTTC	CGG	CGG	CGG

Db 1092 CAGCAGAGCGGTAAGAAAGACCGACTGGATTTCACCAGGCGCTGAAAGGTTACTTC 1151

Qy 1373 AAAATGAATGTGGCAGCAGCAATAATGACAGTCACTGTGTGCTTTTGTCTTCAGTAC 1432

Db 1152 AAAATGAATGTGGCAGCAGCAATAATGACAGTCACTGTGTGCTTTTGTCTTCAGTAC 1211

Qy 1433 TGCAAGCATCTGGCCCTGTCTCAGCCATTTCAGCTTCCACCAGCAGACATGCCCAAACTT 1492

Db 1212 TGCAAGCATCTGGCCCTGTCTCAGCCATTTCAGCTTCCACCAGCAGACATGCCCAAACTT 1271

Qy 1493 CGTCGGCAGATCTACAGAGAGCTGTGTCACGTGCAAACTCACTGTGTGAGCTCTGATCCCC 1552

Db 1272 CGTCGGCAGATCTACAGAGAGCTGTGTCACGTGCAAACTCACTGTGTGAGCTCTGATCCCC 1331

Qy 1553 AGACCCCAAGCCCAATAATGGAGGAGACATGGGAGTCCCTTCCCAAGAACTCCAGT 1612

Db 1332 AGACCCCAAGCCCAATAATGGAGGAGACATGGGAGTCCCTTCCCAAGAACTCCAGT 1391

Qy 1613 TCCTTTCTCTCTTGGCTCTTCCACACTCACTTCCCTTTGGTTTTTTCATATTTAAATGTTT 1672

Db 1392 TCCTTTCTCTCTTGGCTCTTCCACACTCACTTCCCTTTGGTTTTTTCATATTTAAATGTTT 1451

Qy 1673 CAATTTCTGTATTTTTTTTCTTTGAGAGAACTCTTGTGATTTCTGATGTGAGGGGGT 1732

Db 1452 CAATTTCTGTATTTTTTTTCTTTGAGAGAACTCTTGTGATTTCTGATGTGAGGGGGT 1511

Qy 1733 GGCTACAGAAAAGCCCTTTCTCTCTGTTTGCAGGGAGTGTGGCCCTGTGGCCTGGG 1792

Db 1512 GGCTACAGAAAAGCCCTTTCTCTCTGTTTGCAGGGAGTGTGGCCCTGTGGCCTGGG 1571

Qy 1793 TGGAGCAGTCACTCCCTTCCCTTCCCTGAGGAGCAGGAAATCACTGTGTGGGGTGGT 1852

Db 1572 TGGAGCAGTCACTCCCTTCCCTTCCCTGAGGAGCAGGAAATCACTGTGTGGGGTGGT 1631

Qy 1853 GGGCGGACATAGGATCACTGCTCCAGATCTTCAAACTTTTATATATATATATATATA 1912

Db 1632 GGGCGGACATAGGATCACTGCTCCAGATCTTCAAACTTTTATATATATATATATATA 1673

Qy 1913 TAT 1972

Db 1674 --TAT 1731

Qy 1732 GCTCTGCTCAATAAG 1988

Db 1732 GCTCTGCTCAATAAG 1747

RESULT 6

AY008764

LOCUS Mus musculus sentrin/SUMO-specific protease (SEN3) mRNA, complete cds. 2174 bp mRNA linear ROD 21-NOV-2000

DEFINITION

ACCESSION AY008764

VERSION AY008764.1 GI:11245812

KEYWORDS

SOURCE Mus musculus.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2174)

AUTHORS Yeh,E.T., Gong,L. and Kamitani,T.

TITLE Ubiquitin-like proteins: new wines in new bottles

JOURNAL Gene 248 (1-2), 1-14 (2000)

MEDLINE 20267842

PUBMED 10806345

REFERENCE 2 (bases 1 to 2174)

AUTHORS Gong,L. and Yeh,E.T.

TITLE Direct Submission

JOURNAL Submitted (04-OCT-2000) Institute of Molecular Medicine, University of Texas-Houston, 2121 W. Holcombe Blvd. 928, Houston, TX 77030, USA

FEATURES

source

1. .2174

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/chromosome="11"

1. .2174

/gene="SEN3"

52. .1758

/gene="SEN3"

/codon\_start=1

/product="sentrin/SUMO-specific protease"

/protein\_id="AAG33253.1"

/db\_xref="GI:11245813"

/translation="MKETIQGTGSMGPEPPPGTYSNPRRLRLPPLPKPKLKGSGG

GGPDPSGTTPTRLPAPRPSFDASAESEEEDEEEVAANRLPPRMQLGAS

QSRALRPSHRKTCQRRRRAMRFAOMLLYSKSTSLTFHWKLMGRGRRLNAHPKN

HUSPQEGGATPQVPCCRPDRGLPPRLGLLALMAEDGMRGSPFPVSGPMEED

GLURTPKSLDPSGLLCTLPNGFGLSGEGERSLAPDASILISNVCSPGDHVAQ

EKFQSDGLIAEADRTKAGHSPREEHVTCVQSILDEFLTQYTGSLTIDLETV

EKLIEDTFOQEFSTPSRKSLVQLIQSYRMPGNAMVGRFVSKVKTWMDLGTLYG

ONWLNQVNMVGDLYMDTVPEKVHFNFSFYDKLRTKGYDGVKRWTKNVDIENKELL

LIPILHLEVHMSLISVDVRRRTITYFDSPRTLAPRCPKIAKYLOAEAVKRDLDHFQG

WKGYFKMNVARQNNDSGAFVLQYCKHLALSQPFSTQODMPKLRQIYKELCHCKL

TV"

BASE COUNT 513 a 585 c 570 g 506 t

ORIGIN

Query Match 73.3%; Score 1459.2; DB 10; Length 2174;

Best Local Similarity 87.5%; Pred. No. 0;

Matches 1684; Conservative 0; Mismatches 213; Indels 27; Gaps 7;

Qy 1 GCCTCCCTGTCGCCCGCACCTCTTTTGTATGCCTCAGCAAGTGAAGAGGAGGAAGAAGAGG 60

Db 230 GCCTCCCTGCCCGCGGCATCTTTTGTATGCCTCAGCTAGTGAAGAGGAGGAAGAAG 289

Qy 61 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAAGTGGCAGCTTGGAGGCTGCCCCCAAGAT 120

Db 290 AGGAG-----GAAGATGAGGAGGAAGTAGCAGCTTGGAGGCTACCCCTAGGT 337

Qy 121 GGAGTCAGCTGGGAACCTCCCGAGCGGCCCGCTTCCCGCCCGCACATCATCGAAAAACCT 180

Db 338 GGGGCCAACTGGGGGGCTCCCGAGCGCTCTCGAGCTTCCGACCTCTCATAGAAAAACCT 397

Qy 181 GCTCACAGCGCCCGCGCGAGCCATGAGAGCCTTCCGAGTCTGCTCTACTCAAAAAGCA 240

Db 398 GCTCACAGCGCCCGCGCGAGCCATGAGAGCCTTCCAGATGCTCTACTCAAAAAGCA 457

Qy 241 CCTCGCTGACATTCCTACTGGAAAGCTTTGGGGGGCCACCGGGGGCGGGGGGGCCCTCG 300

Db 458 CCTCGCTGACATTCCTACTGGAAAGCTTTGGGGGGCCACCGAGGCGCGCGGAAACCTCG 517

Qy 301 CACACCCCAAGAACCATCTTTTACCCCGAGGGGTGCGAGCCACAGGTGCCATCCC 360

Db 518 CGCACCCCAAGAACCATCTCTCACCCCGAGGGGTGCAACGCCACAGGTGCCATCAC 577

Qy 361 CCTGTTGCTGCTTTTGTAGTCCCGCGGGGCGACCTTCCACCCCGGCTGGGTCTGCTAGGTG 420

Db 578 CTTGCTGCTGCTTTTGTAGTCCCGCGGGGCTACCCCGACCCCGCTGGGTCTGCTAGGTG 637

Qy 421 CTTCTATGCTGAGGATGGGGTGAAGGGTCTCCACAGTGCCTCTGGGGCCCCCCTATGG 480

Db 638 CTTCTATGCGAGAGATGGGATGAGAGGGTCTCCACAGTGCCTCTGGGGCCCCCCTATGG 697

Qy 481 AGGAAGATGGACTCAGTGGACTCCCAAGTCTCTCTGGACCTGACTCGGGGCTCCTTT 540

Db 698 AGGAAGATGGACTAAGTGGAGTGGACCCCAAGTCTCTCTGGACCTGACTCGGGGCTCCTCT 757

Qy 541 CATGTACTCTGCCCAACAGGTTTTTGGGGGCAATCTTGGGCCAGAGGAGGAGCGCA-CTTGG 599

Db 758 CATGTACTCTACCAATAGTTTTTGGGGGACTGTCTGTCTCTGAAGGAGGAGCGCAGCTGG 817

Qy 600 CACCCCGCTGATGCCAGATCCTCATCAGCAANTGTGTGACAGCATCGGGGACCATGTGGCCC 659

Db 818 CACCCCGCTGATGCCAGCATCTCATCAGTAATGTGTGACAGCATTTGAGAGACAGCTGGGCTC 877

QY	660	AGGAGCTTTTTCAGGGCTCAGATTTTGGGCATCGGCAGAGAGCGCAGAGAGCCCTGGGGAGA	719
Db	878	AGGAACCTTTTTCAGAGCTCTGACTTTGGGCATTTGCAGAAAGGCGAGATCGGACTGGGAGA	937
QY	720	AAGCCGGCCAGCACAGCCCCCTCGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGG	779
Db	938	AAGCTGGCCAGCATAGCCCCCTCGGAGAGAACATGTGACCTGTGTGCAAGATATCTTAG	997
QY	780	AGAAATTCCTTCAAACGATANGCAGCCCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA	839
Db	998	ATGAATTCCTTCAAACCTTATGCGACGCTCATCCCTCTCAGCACTGATGAAGATTGTAGAGA	1057
QY	840	AGCTGGAGACATTTTCAGCAGGAGTTTCCACCCCTTCAGGAAGGGCCTGGTGTTC	899
Db	1058	AGTTGAGGACATTTTCAGCAGGAGTTCTCTACACCTCAAGGAAGAGTCTGGTACTAC	1117
QY	900	AGCTGATCCAGTCTTTACAGGGGATGCCAGGCAATGCCATGGTCAGGGGCTTCGAGCTGG	959
Db	1118	AGCTGATCCAGTCTTATCAGCGGATGCCAGCAACGCTATGTTGAAGGGGCTTCGGGTAT	1177
QY	960	CTTATAAGCGGCACGTGCTGACCATGATGACTTGGGACCTTGTATGGACAGAACTTGGC	1019
Db	1178	CTTATAAGCGACACGTGCTACCATGATGACTTTGGGTACCTTATATGGACAGAACTTGGC	1237
QY	1020	TCAATGACCAGGTGATGAACATGATGAGACCTTGGTCATCGACACACAGTCCCTGAAAAGG	1079
Db	1238	TCAATGACCAGGTGATGAACATGATGAGAGTCTGGTCATGGACACAGTCCCTGAAAAGG	1297
QY	1080	TGCATTTCTTCAATAGTTTCTTATGATAAACTCCGTACCAGGGGTATGATGGGTGA	1139
Db	1298	TGCATTTCTTCAACAGCTTCTTATGATAAACTCCGTACCAGGGTATGATGGGGTAA	1357
QY	1140	AAAGTGGACAAAAACGTGGACATCTTCAATAAGGAGTACTGCTAATCCCCATCAACC	1199
Db	1358	AGAGTGGACAAAAATTTGGACATCTTCAATAAGGAATTAAGTCTAATCCCCATCCATC	1417
QY	1200	TGGAGTGCATTGGTCCCTCATCTCTGTGTGATGTAGCGCAGCACCATCACCTATTTTG	1259
Db	1418	TGGAGTGCATGTGTCCCTTATCTCAGTTGATGTAAGCGAGCTACCATCACCTATTTTG	1477
QY	1260	ACTCGACGGTACCCTAAACCGCCGCTCCCTTAAGCATATTGCCAAGTATCTACAGCGAG	1319
Db	1478	ACTCCAGCGAACTCTAAATCGCCGCTCCCTTAAGCATATTGCCAAGTATCTACAGCGAG	1537
QY	1320	AGCGGTAAAGAAAGACCGACTGGATTTCCACAGGGGCTGAAAGTTTACTTCAAAATGA	1379
Db	1538	AGCGAGTCAAAAAGACCGACTGGACTTCCATCAGGGGCTGAAAGTTTACTTCAAAATGA	1597
QY	1380	ATGTGCCAGGCAGAAATATGACAGTGAATGATGAGTGGTCTTTGCTGCTGCAGTACTGCAAGC	1439
Db	1598	ATGTGCCAGGCAGAAATATGACAGTGAATGATGAGTGGTCTTTGATTAAGTACTGCAAGC	1657
QY	1440	ATCTGGCCCTGTCTCAGGCATTCAGCTTTCACCCAGCAGGACATGCCCAAACTTCGTCCGC	1499
Db	1658	ACTGGCCCTGTCTCAGGCATTCAGCTTTCACCCAGCAGGACATGCCCAAACTTCGTCCGC	1717
QY	1500	AGATCTAAGAGAGTGTGTCACTGCAAACTCACTGTGTGAGCCTCGTACCCAGACCCC	1559
Db	1718	AGATCTAAGAGAGTGTGTCACTGCAAACTCACTGTGTGAGCCTCACTACCCAGACCCC	1777
QY	1560	AAGCCCATAAATGGGAAGGAGACATGGAGTCCCTTCCCAAGAACTCCAGTTCCTTC	1619
Db	1778	TAGCCCATCAGTGGCAGG--GGAAGGAGGACCTTCCCAAGAACTCTAGTTCCCT---	1832
QY	1620	CTCTCTTGCTCTTCCCATCTACTTCCCTTTTGGTTTTTCATATTAAATGTTCCTTCAATTC	1679
Db	1833	-----TTTTTGGTTCTCTACCCAAATTCCTTTGGATTTTCATATTAAATGTTCCTTCAATTC	1886
QY	1680	TGTAATTTTTTTTTTCTTTGAGAGAAATACTGTTGATTTCTGATGTGACAGGGGGTGGCTACA	1739
Db	1887	TGTAATTTTTTTTTTCTTTGCTGCGAGTATATGATTTCTGATGGAAGGAATGGCCACA	1946

[illegible]









Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 159951 bases at least Q40  
Consensus quality: 161126 bases at least Q30  
Consensus quality: 161900 bases at least Q20  
Insert size: 185000; agarose-fp  
Quality coverage: 7.39x in Q20 bases; sum-of-contigs  
Quality coverage: 8.34x in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 3137: contig of 3137 bp in length  
\* 3138 3237: gap of unknown length  
\* 3238 13872: contig of 10635 bp in length  
\* 13873 13972: gap of unknown length  
\* 13973 19270: contig of 5298 bp in length  
\* 19271 19370: gap of unknown length  
\* 19371 30584: contig of 11214 bp in length  
\* 30585 30684: gap of unknown length  
\* 30685 45302: contig of 14618 bp in length  
\* 45303 45402: gap of unknown length  
\* 45403 76670: contig of 31268 bp in length  
\* 76671 76770: gap of unknown length  
\* 76771 164504: contig of 87734 bp in length.  
FEATURES  
source  
1. .164504  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="RP43-149M23"  
/clone\_lib="RP43"  
1. .3137  
/note="assembly\_fragment"  
3238. .13872  
/note="assembly\_fragment"  
13973. .19270  
/note="assembly\_fragment"  
19371. .30584  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
30685. .45302  
/note="assembly\_fragment"  
45403. .76670  
/note="assembly\_fragment"  
76771. .164504  
/note="assembly\_fragment"  
BASE COUNT 38842 a 42831 c 41755 g 40472 t 604 others  
ORIGIN  
Query Match 26.6%; Score 530.4; DB 2; Length 164504;  
Best Local Similarity 99.8%; Pred. No. 2.2e-119;  
Matches 531; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GCCTCCCTGTCGCCCGACCCCTCTTTTGATGCCCTCAGCAGTGAAGAGGAGGAAGAGG 60  
Db 25116 GCCTCCCTGTCGCCCGACCCCTCTTTTGATGCCCTCAGCAGTGAAGAGGAGGAAGAGG 25057  
QY 61 AGGAGGAGGAGGATGAAGATGAAGAGAGAGAGAGTGGCAGCTTGGAGCTGCCCGCCCAAGAT 120  
Db 25056 AGGAGGAGGAGGATGAAGATGAAGAGAGAGAGTGGCAGCTTGGAGCTGCCCGCCCAAGAT 24997  
QY 121 GGAGTCAGCTGGGAACCTCCACGCGCCCGCCCTTCCCGCCCGCCCTCATCGAAAAACCT 180  
Db 24996 GGAGTCAGCTGGGAACCTCCACGCGCCCGCCCTTCCCGCCCGCCCTCATCGAAAAACCT 24937

QY 181 GCTCACAGCGCCCGCCCGAGCCATGAGAGCCCTCCGGATGCTGCTTACTCTAAAAAGCA 240  
Db 24936 GCTCACAGCGCCCGCCCGAGCCATGAGAGCCCTCCGGATGCTGCTTACTCTAAAAAGCA 24877  
QY 241 CCTCGCTGACATTCACACTGGAAGCTTTTGGGGGGCCACCGGGCCCGCGGGGCTCG 300  
Db 24876 CCTCGCTGACATTCACACTGGAAGCTTTTGGGGGGCCACCGGGCCCGCGGGGCTCG 24817  
QY 301 CACACCCCAAGAACCACATCTTTTACCCACAGCAAGGGGGTGGAGCCACACAGGTGCCATCCC 360  
Db 24816 CACACCCCAAGAACCACATCTTTTACCCACAGGAAGGGGTGGAGCCACACAGGTGCCATCCC 24757  
QY 361 CTGTGTTGCTGTTTGTACTCCCCCGGGGGCCACCTCCACCCCGGGCTGGTCTGCTAGGTG 420  
Db 24756 CTGTGTTGCTGTTTGTACTCCCCCGGGGGCCACCTCCACCCCGGGCTGGTCTGCTAGGTG 24697  
QY 421 CTCTCATGCTGAGGATGGGTGAGAGGCTCTCCACCACTGCTGCGCCCTTGGGCCCGCCATGG 480  
Db 24696 CTCTCATGCTGAGGATGGGTGAGAGGCTCTCCACCACTGCTGCGCCCTTGGGCCCGCCATGG 24637  
QY 481 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCCTCTGGACCCCTGACTCGGG 532  
Db 24636 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCCTCTGGACCCCTGACTCGGG 24585  
RESULT 10  
AC127470 215795 bp DNA linear HTG 17-JUL-2002  
LOCUS Pan troglodytes clone RP43-145D13, WORKING DRAFT SEQUENCE, 15  
DEFINITION unordered pieces.  
AC127470  
VERSION AC127470.1 GI:21886866  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE chimpanzee  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
1 (bases 1 to 215795)  
Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S. M.,  
Benjamin, B., Blakesley, R. W., Bouffard, G. G., Breen, K., Brinkley, C.,  
Brooks, S., Dietrich, N. L., Granite, S., Guan, X., Gupta, J.,  
Haghighi, P., Hansen, N., Ho, S.-L., Idol, J. R., Karlins, E., Laric, P.,  
Lee-Lin, S.-Q., Legaspi, R., Maduro, Q. L., Maduro, V. B.,  
Margulies, E. H., Mastello, C., Maskeri, B., Mastrian, S. D.,  
McCloskey, J. C., McDowell, J., Paguirigan, C., Pearson, R.,  
Portnoy, M. E., Prasad, A., Schueler, M. G., Stantropop, S., Thomas, J. W.,  
Thomas, P. J., Touchman, J. W., Tsurgan, C., Vogt, J. L., Walker, M. A.,  
Wetherby, K. D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E. D.  
NISC Comparative Sequencing Initiative  
Unpublished  
2 (bases 1 to 215795)  
Green, E. D.  
Direct Submission  
Submitted (17-JUL-2002) NIH Intramural Sequencing Center, 8717  
Grovemont Circle, Gaithersburg, MD 20877, USA  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: <http://www.nisc.nih.gov>  
Contact: [nisc\\_zooenhgri.nih.gov](mailto:nisc_zooenhgri.nih.gov)  
----- Project Information  
Center project name: cms  
Center clone name: 145D13  
----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 203929 bases at least Q40  
Consensus quality: 206865 bases at least Q30  
Consensus quality: 209181 bases at least Q20  
Insert size: 190000; agarose-fp  
Insert size: 214395; sum-of-contigs  
Quality coverage: 7.30x in Q20 bases; agarose-fp





REFERENCE 1 (bases 1 to 5318)  
AUTHORS Graves,D.R., Thomsen,L., Catchpole,I.R. and Ford,M.J.  
TITLE Dna constructs based on the eif4a gene promoter  
JOURNAL Patent: WO 0102594-A 40 11-JAN-2001;  
GLAXO GROUP LIMITED (GB)  
FEATURES  
source Location/Qualifiers  
1..5318  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 1327 a 1210 c 1423 g 1350 t 8 others  
ORIGIN  
Query Match 24.4%; Score 486.4; DB 6; Length 5318;  
Best Local Similarity 94.9%; Pred. No. 9.8e-109;  
Matches 544; Conservative 1; Mismatches 17; Indels 11; Gaps 4;  
QY 1427 CAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCAGCTTCACCCAGGACATGCC 1486  
|||||  
Db 3827 CAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCAGCTTCACCCAGGACATGCC 3886  
QY 1487 AAACCTTCGTGGCAGATCTACAAAGAGCTGTGTCTACTGCAAACTCACGTGTGAGCCTCG 1546  
|||||  
Db 3887 AAACCTTCGTGGCAGATCTACAAAGAGCTGTGTCTACTGCAAACTCACGTGTGAGCCTCG 3946  
QY 1547 TACCCAGACCCCAAGCCCAATAAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAAAC 1606  
Db 3947 TACCCAGACCCCAAGCCCAATAAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAAAC 4006  
QY 1607 TCCAGTTCCTTTCCTCTCTTGCCTCTTCCCACTACTTCCCTTTGGTTTTCATATTAA 1666  
|||||  
Db 4007 TCCAGTTCCTTTCCTCTCTTGCCTCTTCCCACTACTTCCCTTTGGTTTTCATATTAA 4066  
QY 1667 ATGTTTCAATTTC-TGTATTTTTTTTCTTGAGAGAATACTTGTATTTCTGATGTC 1725  
|||||  
Db 4067 ATGTTTCAATTTC-TGTATTTTTTTTCTTGAGAGAATACTTGTATTTCTGATGTC 4126  
QY 1726 AGGGGTGGCTACAGAAAGCCCTTCTCTCTGTTTGAGGGAGTGTGGCCCTGTG 1785  
|||||  
Db 4127 AGGGGATGGCTACAGAAAGCCCTTCTCTCTGTTTGAGGGAGTGTGGCCCTGTG 4186  
QY 1786 G-CCTGGTGGAGCAGTCACTCTCCCTTCCCTTCCCTGGAGGAGCAGGAATCAGTGCT- 1843  
|||||  
Db 4187 GCCCTGGTGGAGCAGTCACTCTCCCTTCCCTTCCCTGGAGGAGCAGGAATCAGTGATG 4246  
QY 1844 GGGGTGGTGGCGGACAAATAGATCACTGCTGCTGCCAGATCTTCAAACTTTTATATAT 1903  
|||||  
Db 4247 GGGGNGGGGGGGGACAAATAGATNACAGCCCGCCAGATNCAATATATATATATAT 4306  
QY 1904 AT 1957  
Db 4307 AT 4366  
QY 1958 --AAATGCCAGGTCTGCTGTGGTCAATAAAG 1988  
|||||  
Db 4367 AAAAATGCCAGGTCTGCTGTGGTCAATAAAG 4399  
RESULT 13  
AF175325  
LOCUS AF175325 6732 bp DNA linear PRI 14-FEB-2000  
DEFINITION Homo sapiens eukaryotic initiation factor 4AI (EIF4AI), gene,  
partial cds.  
ACCESSION AF175325  
VERSION AF175325.1 GI:6969997  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.  
REFERENCE 1 (Bases 1 to 6732)  
AUTHORS Quinn,C.M., Wiles,A.P., El-Shanawany,T., Catchpole,I., Alnadaf,T.,  
Ford,M.J., Gordon,S. and Greaves,D.R.  
TITLE The human eukaryotic initiation factor 4AI gene (EIF4AI) contains  
multiple regulatory elements that direct high-level reporter gene  
expression in mammalian cell lines  
JOURNAL Genomics 62 (3), 468-476 (1999)  
MEDLINE 20112761  
PUBMED 10644445  
REFERENCE 2 (bases 1 to 6732)  
AUTHORS Greaves,D.R. and Catchpole,I.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Sir William Dunn School of Pathology,  
University of Oxford, South Parks Road, Oxford OX1 3RE, UK  
FEATURES  
source Location/Qualifiers  
1..6732  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17p13"  
/clone="cosCD68C1"  
repeat\_region 656..1019  
/rpt\_family="Alu I"  
/rpt\_type-dispersed  
repeat\_region 1521..1823  
/rpt\_family="Alu I"  
/rpt\_type-dispersed  
misc\_feature 3868..4120  
/note="similar to murine eIF-4AI gene"  
repeat\_region 4259..4334  
/rpt\_type-tandem  
/rpt\_unit-at  
misc\_feature 4773..5193  
/note="similar to murine eIF-4AI gene"  
gene 5249..>6732  
/gene="EIF4AI"  
mRNA join(5249..5286,6681..>6732)  
/gene="EIF4AI"  
exon /product="eukaryotic initiation factor 4AI"  
5249..5286  
/gene="EIF4AI"  
/number=1  
CDS join(5264..5286,6681..>6732)  
/gene="EIF4AI"  
/codon\_start=1  
/note="ribosome component"  
/product="eukaryotic initiation factor 4AI"  
/protein\_id="AAF34151.1"  
/db\_xref="GI:6969998"  
exon /translation="MSASQDSRSRDNGPDGMEPEGVIE"  
6681..6732  
/gene="EIF4AI"  
/number=2  
BASE COUNT 1555 a 1597 c 1978 g 1594 t 8 others  
ORIGIN  
Query Match 24.4%; Score 486.4; DB 9; Length 6732;  
Best Local Similarity 94.9%; Pred. No. 1e-108;  
Matches 544; Conservative 1; Mismatches 17; Indels 11; Gaps 4;  
QY 1427 CAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCAGCTTCACCCAGGACATGCC 1486  
|||||  
Db 3795 CAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCAGCTTCACCCAGGACATGCC 3854  
QY 1487 AAACCTTCGTGGCAGATCTACAAAGGAGCTGTGTCTACTGCAAACTCACGTGTGAGCCTCG 1546  
|||||  
Db 3855 AAACCTTCGTGGCAGATCTACAAAGGAGCTGTGTCTACTGCAAACTCACGTGTGAGCCTCG 3914  
QY 1547 TACCCAGACCCCAAGCCCAATAAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAAAC 1606  
|||||  
Db 3915 TACCCAGACCCCAAGCCCAATAAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAAAC 3974  
QY 1607 TCCAGTTCCTTTCCTCTCTTGCCTCTTCCCACTACTTCCCTTTGGTTTTCATATTAA 1666  
Db 3975 TCCAGTTCCTTTCCTCTCTTGCCTCTTCCCACTACTTCCCTTTGGTTTTCATATTAA 4034  
QY 1667 ATGTTTCAATTTC-TGTATTTTTTTTCTTGAGAGAATACTTGTATTTCTGATGTC 1725



Db 136003 GGGGTGGCTACAGAAACCCCTTCTTCCTCTGTTTCAGCGGAGTGTGCCCTGTGG 136062  
QY 1787 CTTGGTGGAGCAGTCACTCCCTCCCTCCCTGTCAGGAGCAGGAAATCAGTGTGG 1846  
Db 136063 CTTGGTGGAGCAGTCACTCCCTCCCTGTCAGGAGCAGGAAATCAGTGTGG 136122  
QY 1847 GGTGTGGCGGACATAGATCACTGCTCCAGATCTTCAACTTTTATATATATA 1906  
Db 136123 GGTGTGGCGGACATAGATCACTGCTCCAGATCTTCAACTTTT-TTTTTTTTA 136181  
QY 1907 TATATATATATATATAT 1925  
Db 136182 TATATATATATATATAAT 136200

RESULT 15  
AC126925/c  
LOCUS 161428 bp DNA linear HTG 10-JUL-2002  
DEFINITION Canis familiaris clone RP81-332E11, WORKING DRAFT SEQUENCE, 15  
unordered pieces.  
AC126925  
VERSION AC126925.1 GI:21724102  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE dog.  
ORGANISM Canis familiaris

REFERENCE 1 (bases 1 to 161428)  
AUTHORS Akter.N., Antonellis.A., Ayele.K., Beckstrom-Sternberg.S.M.,  
Benjamin.B., Blakesley.R.W., Bouffard.G.G., Breen.K., Brinkley.C.,  
Brooks.S., Dietrich.N.L., Granite.S., Guan.X., Gupta.J.,  
Haghighi.P., Hansen.N., Ho.S.-L., Idol.J.R., Karlins.E., Laric.P.,  
Lee-Lin.S.-Q., Legaspi.R., Maduro.Q.L., Maduro.V.B.,  
Marquies.E.H., Mastello.C., Maskeri.B., Mastrian.S.D.,  
McCloskey.J.C., McDowell.J., Paguirigan.C., Pearson.R.,  
Portnoy.M.E., Prasad.A., Schueler.M.G., Stantripop.S., Thomas.J.W.,  
Thomas.P.J., Touchman.J.W., Tsurgeon.C., Vogt.J.L., Walker.M.A.,  
Wetherby.K.D., Wiggins.L., Young.A., Zhang.L.-H. and Green,E.D.  
NISC Comparative Sequencing Initiative  
Unpublished

2 (bases 1 to 161428)  
Green,E.D.  
Direct Submission  
Submitted (10-JUL-2002) NIH Intramural Sequencing Center, 8717  
Grovetown Circle, Gaithersburg, MD 20877, USA  
----- Genom Center Sequencing Center

Center code: NISC  
Center site: http://www.nisc.nih.gov  
Web site: http://www.nisc.nih.gov  
Contact: nisc.zoo@nhgri.nih.gov  
----- Project Information  
Center project name: cwp  
Center clone name: 332E11

----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 148712 bases at least Q40  
Consensus quality: 151471 bases at least Q30  
Consensus quality: 153125 bases at least Q20  
Insert size: 152000; agarose-fp  
Insert size: 160028; sum-of-contigs  
Quality coverage: 5.36x in Q20 bases; agarose-fp  
Quality coverage: 5.10x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 4734: contig of 4734 bp in length  
\* 4835: gap of unknown length  
\* 4835: contig of 2491 bp in length  
\* 7326: gap of unknown length  
\* 7426: contig of 4836 bp in length  
\* 12261: gap of unknown length  
\* 12362: contig of 6670 bp in length  
\* 19031: gap of unknown length  
\* 19131: contig of 8718 bp in length  
\* 27849: gap of unknown length  
\* 27850: contig of 3852 bp in length  
\* 27950: gap of unknown length  
\* 31802: contig of 8880 bp in length  
\* 31902: gap of unknown length  
\* 40782: contig of 8902 bp in length  
\* 40882: gap of unknown length  
\* 49784: contig of 9267 bp in length  
\* 49884: gap of unknown length  
\* 59151: contig of 9429 bp in length  
\* 59251: gap of unknown length  
\* 68680: contig of 10205 bp in length  
\* 68780: gap of unknown length  
\* 78985: contig of 14362 bp in length  
\* 79085: gap of unknown length  
\* 93447: contig of 8629 bp in length  
\* 93547: gap of unknown length  
\* 102176: contig of 29410 bp in length  
\* 102276: gap of unknown length  
\* 131685: contig of 29410 bp in length  
\* 131686: gap of unknown length  
\* 131786: contig of 29643 bp in length.

FEATURES  
Location/Qualifiers  
1..161428  
/organism="Canis familiaris"  
/db\_xref="taxon:9615"  
/clone\_lib="RP81-332E11"  
/clone\_lib="RP81"

misc\_feature 1..4734  
/note="assembly\_fragment"  
misc\_feature 4835..7325  
/note="assembly\_fragment"  
misc\_feature 7426..12261  
/note="assembly\_fragment"  
misc\_feature 12362..19031  
/note="assembly\_fragment"  
misc\_feature 19132..27849  
/note="assembly\_fragment"  
misc\_feature 27950..31801  
/note="assembly\_fragment"  
misc\_feature 31902..40781  
/note="assembly\_fragment"  
misc\_feature 40882..49783  
/note="assembly\_fragment"  
misc\_feature 49884..59150  
/note="assembly\_fragment"  
misc\_feature 59251..68679  
/note="assembly\_fragment"  
misc\_feature 68780..78984  
/note="assembly\_fragment"  
misc\_feature 79085..93446  
/note="assembly\_fragment"  
misc\_feature 93547..102175  
/note="assembly\_fragment"  
misc\_feature clone\_end:SP6  
vector\_side:right  
misc\_feature 102276..131685  
/note="assembly\_fragment"  
misc\_feature 131786..161428  
/note="assembly\_fragment"  
BASE COUNT 39586 a 41863 c 41233 g 37340 t 1406 others  
ORIGIN

Query Match 23.8%; Score 474.4; DB 2; Length 161428;  
Best Local Similarity 93.2%; Pred. NO. 1.3e-105;  
Matches 496; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 GCCTCCCTGTCCCCGACCCCTCTTTTGTATGCCTCAGCAAGTGAAGAGAGAGAGAGAGG 60  
|||||  
Db 88615 GCCTCCCTGTCCCTCGGCCCTCTTTTGTATGCCTCAGCTAGTGAAGAGAGAGAGAGAGG 88556  
QY 61 AGGAGGAGGAGATGAAGATGAAGAGAGAGAGAGAGAGTGGCAGCTTGGAGGCTGCCCCCAAGAT 120  
|||||  
Db 88555 AGGAAGATGAGGAGAGATGAGGAAGAGAGAGTGGCAGCTTGGAGGCTGCCCTCCAAGAT 88496  
QY 121 GGAGTCAAGCTGGGAACCTCCAGCGGCCCGCCCTTCCGCCCCCACTCATCGAAAAACCT 180  
|||||  
Db 88495 GGGGTCAAGCTGGGGCCCTCCAGCGACCTCGCCCTCCCGGCCCTACTCATCGAAAAACCT 88436  
QY 181 GCTCACAGCGCCGCCCGAGCCATGAGAGCCTTCCGGATGCTGCTCTACTCAAAAAAGCA 240  
|||||  
Db 88435 GCTCACAGCGCCGCCCGAGCCAAAGAGCTTTCGGATGCTGTTCTACCCAAAAAGCA 88376  
QY 241 CCTCGCTGACATTCACATTCGAAGCTTTGGGGGCCACCGGGGCCGGGGGCCCTCG 300  
|||||  
Db 88375 CCTCGTTGACATTCACATTCGAAGCTTTGGGGGCCACCGGGGCCGGGGGCCCTCG 88316  
QY 301 CACACCCCAAGAACATCTTTTCACCCCAGCAAGGGGTGGACGCCACAGGTGCCATCCC 360  
|||||  
Db 88315 CACACCCCAAGAACATCTTTTCACCCCAGCAAGGGGTGGACGCCACAGGTGCCATCTC 88256  
QY 361 CCTGTTGCTGTTTGTACTCCCGCCGGGGGCCACTCCACCCCGGCTGGGTCTGCTAGGTG 420  
|||||  
Db 88255 CCTGCTGCTGTTTGTACTCCCGCTCGGGGGCCACTCCACCCCGGCTGGGTCTGCTAGGTG 88196  
QY 421 CTCTCATGGCTGAGATGGGTGAGAGGGTCTCCACCAAGTCCCTCTGGGCCCCCAATGG 480  
|||||  
Db 88195 CTCTCATGGCTGAGATGGGTGAGAGGGTCTCCACCAATGCTCTCTGGGCCCCCAATGG 88136  
QY 481 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCCTCTGGACCTGACTCGGG 532  
|||||  
Db 88135 AGGAAGATGGATTAAGGTGGACTCCAAAGTCTCCTCTGGATCTGACTCTGG 88084

Search completed: December 16, 2002, 17:29:13  
Job time : 6622 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2002, 21:18:11 ; Search time 2422 Seconds  
(without alignments)  
2220.025 Million cell updates/sec

Title: US-09-848-852A-3  
Perfect score: 1772  
Sequence: 1 MYSAQRFWGTIARRGAHLA.....DMPKRRQIYKELCHCKLTV 332

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=fragnet\_p2n\_model -DEV=xlp  
-Q=/cqn2\_1/USPTO\_spool/US09848852/runat\_16122002\_132040\_19159/app\_query.fasta\_1.519  
-DB=EST -QFMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09848852.qcgn\_1.1.2024.@runat\_16122002\_132040\_19159 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_estl:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1391	78.5	1071	14	BM806659	BM806659 AGENCOURT
2	1380	77.9	1060	14	BM806502	BM806502 AGENCOURT
3	1250	70.5	804	12	BG870520	BG870520 602791439
4	1172.5	66.2	728	12	BG865703	BG865703 602783877
5	1133	63.9	985	13	BI456217	BI456217 603172996
6	1097	61.9	626	14	BM696661	BM696661 UI-E-DW0-
7	1073	60.6	905	14	BQ898865	BQ898865 AGENCOURT
8	985.5	55.6	1134	14	BM804530	BM804530 AGENCOURT
c 9	979	55.2	667	17	AZ399027	AZ399027 IMO164L10
c 10	975	55.0	616	10	AW916790	AW916790 EST348198
c 11	951	53.7	620	10	AW578424	AW578424 RCI-CT024
c 12	951	53.7	620	10	AW604359	AW604359 RCI-CT024
c 13	929	52.4	572	10	BE077141	BE077141 RCI-CT024
c 14	924	52.1	626	9	AL578220	AL578220 AL578220
c 15	913	51.5	870	13	BI152305	BI152305 602917742
c 16	903	51.0	549	13	BM507574	BM507574 ih4le03.y
c 17	877	49.5	593	10	AW643059	AW643059 cm25f04.w
c 18	865	48.8	819	12	BF139554	BF139554 601785671
c 19	852	48.1	545	17	AQ978128	AQ978128 RPCI-23-3
c 20	834	47.1	504	12	BF954315	BF954315 QV2-NN004
c 21	833	47.0	621	13	BI392544	BI392544 pcpin.pk0
c 22	818	46.2	477	10	AW825381	AW825381 us14e04.y
c 23	812	45.8	817	9	AL581080	AL581080 AL581080
c 24	809	45.7	489	12	BF458035	BF458035 UI-M-BZ1-
c 25	804	45.4	458	14	W63893	W63893 md85h02.r1
c 26	801	45.2	835	13	BI732013	BI732013 603355881
c 27	798	45.0	449	9	AA549588	AA549588 vk76e01.s
c 28	798	45.0	689	12	BF608506	BF608506 MY1-00144
c 29	794	44.8	472	14	BM821282	BM821282 K-EST0090
c 30	794	44.8	605	14	BM829799	BM829799 K-EST0102
c 31	794	44.8	1070	13	BI247525	BI247525 602960160
c 32	789	44.5	1828	11	BC000958	BC000958 Homo sapi
c 33	784	44.2	1836	11	AF335474	AF335474 Homo sapi
c 34	782	44.1	447	9	AI148063	AI148063 qg61g08.r
c 35	782	44.1	468	9	AI261629	AI261629 q230e09.x
c 36	782	44.1	478	14	BM941363	BM941363 UI-M-BZ1-
c 37	776	43.8	599	13	BI017569	BI017569 PK3-ET027
c 38	770	43.5	1099	13	BM55097	BM55097 AGENCOURT
c 39	768	43.3	632	12	BG083248	BG083248 H3086807-
c 40	765	43.2	650	14	BQ261255	BQ261255 fz71g09.x
c 41	759	42.8	665	13	BM185366	BM185366 fv46c09.x
c 42	753	42.5	1158	14	BQ934444	BQ934444 AGENCOURT
c 43	749	42.3	595	14	BQ395166	BQ395166 NISC_bg13
c 44	742	41.9	676	10	AW961869	AW961869 EST373942
c 45	738	41.6	499	12	BF443604	BF443604 261154 MA

ALIGNMENTS

RESULT 1  
BM806659

LOCUS

DEFINITION BM806659 1071 bp mRNA linear EST 05-MAR-2002  
5', mRNA sequence.

ACCESSION BM806659

VERSION BM806659.1 GI:19123482

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLML2258 row: 1 column: 15  
High quality sequence stop: 648.  
Location/Qualifiers

FEATURES  
source

1..1071  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5548958"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site.1: NotI;  
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."  
255 a 264 c 271 g 278 t 3 others

BASE COUNT  
ORIGIN

Alignment Scores:  
Pred. No.: 2,28e-169 Length: 1071  
Score: 1391.00 Matches: 264  
Percent Similarity: 99.25% Conservative: 0  
Best Local Similarity: 99.25% Mismatches: 1  
Query Match: 78.50% Indels: 1  
DB: 14 Gaps: 0

US-09-848-852a-3 (1-332) x BM806659 (1-1071)

QY 68 ArgGluGluHisValThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTrpGly 87  
|||||  
Db 2 CGAGAGGAGCATGTGACCTGGGTACAGAGCATCTTGGAGCAATCTTCAACGATATGGC 61  
QY 88 SerLeuIleProLeuSerThrAspGluValValGluLeuGluAspIlePheGlnGln 107  
|||||  
Db 62 AGCCTCATACCCCTCAGCACTGATGAGTAGTAGAGAAAGCTGGAGGACATTTTCCAGCAG 121  
QY 108 GluPheSerThrProSerArgLysGlyLeuValLeuGluLeuIleGlnSerTrpGlnArg 127  
|||||  
Db 122 GAGTTTCCACCCCTCCAGGAAGGGCGTGTGTTGCAGCTGATCCAGTCTTACCAAGCGG 181  
QY 128 MetProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThr 147  
|||||  
Db 182 ATGCCAGGCAATGCCATGGTGGAGGGCTTCCGAGTGGCTTATAAGCGGCACGTGCTGACC 241  
QY 148 MetAspAspLeuGlyThrLeuTrpGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMet 167  
|||||  
Db 242 ATGGATGACTTGGGGACCTGTGTGGACAGAACTGGCTCAATGACCAAGGTGATGAACATG 301  
QY 168 TyrGlyAspLeuValMetAspThrValProGluLysValHisPheAsnSerPhePhe 187  
|||||  
Db 302 TATGGAGACTGTGTCATGACACAGTCCCTGAAAGAGTGCATTTCTCAATAGTTTCTTC 361  
QY 188 TyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTrpThrLysAsnValAsp 207  
|||||  
Db 362 TATGATAAATCCGTACCAAGGGTTATGATGGGTCAAAAGGTGGACCAAAACGTTGGAC 421  
QY 208 IlePheAsnLysGluLeuLeuLeuProIleHisLeuGluValHisTrpSerLeuIle 227  
|||||  
Db 422 ATCTTCAATAAGAGAGTACTGTCTAAATCCCATCCACTGGAGTGCATTTGGTCCCTCATC 481  
QY 228 SerValAspValArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArg 247  
|||||  
Db 482 TCTGTTGATGTGAGGCGACGACCATCACCTATTTTGGACTCGCAGCGTACCCCTAAACCCG 541

QY 248 ArgCysProLysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAspArgLeu 267  
|||||  
Db 542 CGTGCCTTAAGCATATTGCCAAGTATCTACAGCGAGCGCGTAAAGAACCCGACTG 601  
QY 268 AspPheHisGlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAsnAsnAsp 287  
|||||  
Db 602 GATTTCACACAGCGCTGGAAAGGTTACTTCAAAATGAATGTGGCCAGGAGAGATAATGAC 661  
QY 288 SerAspCysGlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPhe 307  
|||||  
Db 662 AGTCACTGTGGTGCTTTTGTGTTCAGTACTGCAACATCTGGCCCTGTCTCAGCCATTC 721  
QY 308 SerPheThrGlnGlnAspMetProLysLeuArgArgGlnIleTyrLysGluLeuCysHis 327  
|||||  
Db 722 AGCTTCACCCAGCAGCAGCATGCCAAACTTCGTGCGCAGATCTFACAAGGAGCTGTGTCTCAC 781  
QY 328 Cys-LysLeuThrVal 332  
|||||  
Db 782 TGGCAAACTCACTGTG 797

## RESULT 2

BM806502

## LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

```
|||||
Db 40 ATGTACTCTGCCACGGTTTGGGGCAAACTCTGGCCAGAGGGAGCGCAGCTTGGC 99
QY 20 aProProaspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
Db 100 ACCCCCTGATGCAGCATCTCATCAGCAATGTGTCAGCATCGGGGACCATGTGGCCCA 159
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluAraGlyProGlyGluLy 60
Db 160 GGAGCTTTTCAGGCTCAGATTGGGCATGGCAGAGGAGGAGAGCCCTGGGGAGNA 219
QY 60 salaGlyGlnHisSerProLeuAraGluGluHisValThrCysValGlnSerIleLeuAs 80
Db 220 ACCCGGCCAGCACACCCCTCGAGAGAGCATGTGACCTGCTGACAGCATCTTGA 279
QY 80 pGluPheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluValValGluLy 100
Db 280 CGAATTCCTCAACAGTATGGCAGCTCATACCCCTCAGCACTGATGAGGTAGTAGAA 339
QY 100 sLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGl 120
Db 340 GCTGGAGGACATTTCCAGCAGAGTTTCCACCCCTCCAGAGAGGCGCTGCTTGA 399
QY 120 nLeuIleGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheAraValAl 140
Db 400 GCTGATCCAGCTTTACACGCGATGCCAGGCAATGCCATGTGTGAGGGGCTCCGAGTGC 459
QY 140 aTyrglyArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrglyGlnAsnTrpLe 160
Db 460 TTATAAGGGGACGTCGTACCATGGATGACCTGGGGACCTTGTATGGACAGAACTGGCT 519
QY 160 uAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThrValProGluLysVa 180
Db 520 CAATCACCAGGTGATGACATGTATGGAGACCTGTGTCATGGACACAGCCCTGAAAGGT 579
QY 180 lHisPheAsnSerPhePheTyrglyAspLysLeuArgThrLysGlyTyrglyValLy 200
Db 580 GCATTTCTTCAATAGTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGTGAA 639
QY 200 sArgThrThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220
Db 640 AAGGTGGACCAAAACGTGGACATCTCAATAAGAGGTACTGCTAAAT-CCCATCCACT 698
QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrglyPheAs 240
Db 699 GGAGGTGATTTGGTCCCTCATCTCTGTCATGTGAGGCGACGCCATCCTATTGTA 758
QY 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrglyLeuGlnAlaGl 260
Db 759 CTCGAGCGGTACCTAGACCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAGA 818
QY 260 u-AlaValLys-LysAspArgLeu-AspPheHisGlnGly---TrpLysGlyTyrgly 278
Db 819 AGCGCGTTAAGAAACACCGACTGGGATTTCCACCACGGGCTGGGAAAGGGTTACTCAA 878
QY 278 sMetAsn-ValAlaArgGlnAsn-AsnAspSerAsp---CysGlyAlaPhe 293
Db 879 AATGAATGTCGCCAGGCAAGATAATGACACTGACCGGTGGGGGCTTTT 929

RESULT 3
BG870520
LOCUS 602791439F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4922696 5',
DEFINITION BG870520.1 Gr:14221060
ACCESSION BG870520
VERSION BG870520.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 804)
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
```

```
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapps-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILLNL at:
http://image.llnl.gov
Plate: LLAM10842 row: n column: 09
High quality sequence stop: 789.
FEATURES
source
1..804
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4922696"
/lab_host="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site:1:
NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 223 a 195 c 202 g 184 t
ORIGIN
Alignment Scores: 2.99e-151 Length: 804
Pred. No.: 1250.00 Matches: 235
Score: 98.74% Conservative: 1
Percent Similarity: 98.33% Mismatches: 2
Best Local Similarity: 98.33% Indels: 1
Query Match: 70.54% Gaps: 0
DB: 12
US-09-848-852A-3 (1-332) x BG870520 (1-804)
QY 95 AspGluValValGluLysLeuGluAspIlePheGlnGlnPheSerThrProSerArg 114
Db 38 GATGAAGTTGTAGAGAAGTTGGAGGACATTTCCAGCAGGAGTTCTTACACCCCTCAAGG 97
QY 115 LysGlyLeuValLeuGlnLeuIleGlnSerTyrglnArgMetProGlyAsnAlaMetVal 134
Db 98 AAGAGCTGGTACTACAGCTGATCCAGTCTTATCAGCGGATGCCAGCAACCTATGGTA 157
QY 135 ArgGlyPheArgValAlaTyrglyArgHisValLeuThrMetAspAspLeuGlyThrLeu 154
Db 158 AGGGGCTTCCGGGTATCTCTAAGCGCACACGTCCTCACCATGGATGGGTACCTTA 217
QY 155 TyrglyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAsp 174
Db 218 TATGACAGAACTGCTCAATCACCAGGTGATGAACATGTATGGAGATCTGGTCATGGAC 277
QY 175 ThrValProGluLysValHisPheAsnSerPhePheTyrglyAspLysLeuArgThrLys 194
Db 278 ACAGTCCCTGAAAAGTGCATTTCTCAACAGCTTCTTCTATGATAAACTCCGTACCAAG 337
QY 195 GlyTyrglyValLys-ArgThrThrLysAsnValAspIlePheAsnLysGluLeuLeu 214
Db 338 GGTATGATGGGGTAAACGAGGTGGACCAAAAATGTGGACATCTTCAATAAGGAATTA 397
QY 214 uLeuIleProIleHisLeuGluValHisTrpSerLeuIleSerValAspValArgArg 234
Db 398 GCTAATCCCATCCATCTGGAGGTGACCTGGTCCCTTATCTCAGTTGATGTAAAGCGCAG 457
QY 234 gThrIleThrTyrglyAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAl 254
Db 458 TACCATCATCTATTTTGTAGCTCCAGCGAAGCTTAATCGCCGCTGCCCTAAGCATATGC 517
QY 254 aLysTyrglyLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLy 274
Db 518 CAAGTATCTACAGGCAGAGGCGAGTCAAAAAGACCGACTGGAGTTCCTCCATCAGGCTGGA 577
```

```

QY 274 sGlyTyrPhelYsMetAsnValAlaArGInAsnAspSerAspCysGlyAlaPheVa 294
|||||
Db 578 AGGTACTTCARAAATGAATGGCCAGCAGATAATGACAGTGACTGGTGGCTTTGT 637

QY 294 lLeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMe 314
|||||
Db 638 ATTACAGTACTGCAAGCACCTGGCCCTGTCTCAGCCATTACAGCTTCCACCCAGCAGGACAT 697

QY 314 tProLysLeuArgArgGlnIleYrLysGluLeuCysHisCysLysLeuThrVal 332
|||||
Db 698 GCCCAAACTCCGTGTCAGATCTACAGGAAGTGTGTCTGCTGCAAACTGCTG 752

RESULT 4
BG865703 728 bp mRNA linear EST 29-MAY-2001
LOCUS 602783877F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4909938 5',
DEFINITION mRNA sequence.
ACCESSION BG865703
VERSION BG865703.1 GI:14216243
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 728)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10809 row: j column: 19
High quality sequence stop: 728.
FEATURES
source
Location/Qualifiers
1..728
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4909938"
/lab_host="NCI_CGAP_SG2"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 198 a 168 c 186 g 176 t
ORIGIN

Alignment Scores:
Pred. No.: 2,99e-141 Length: 728
Score: 1172.50 Matches: 234
Percent Similarity: 97.53% Conservative: 3
Best Local Similarity: 96.30% Mismatches: 6
Query Match: 66.17% Indels: 3
DB: 12 Gaps: 0

US-09-848-852A-3 (1-332) x BG865703 (1-728)

QY 29 SerAsnValCysSerTleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAspLeu 48
|||||
Db 2 AGTAATGTGTGCAGCATGTGAGACCAACGCTGGCTCAGGAACCTTTTCAGAGCTCTGACTTG 61

QY 49 GlyMetAlaGluAlaGluArgProGlyGluLysAlaGlyClnHisSerProLeuArg 68
|||||
Db 62 GGCATGCA-GAAGAGGACGATCGGCTGGGGAGAAAGCTGGCCAGCATAGCCCCCTCGG 120

```

```

QY 69 GluGluHisValThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGlySer 88
|||||
Db 121 GAGGACATGTGACCTGTGTGCAGAGATATCTTAGATGAATTCCTTCAAACTTATGGCAGC 180

QY 89 LeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGlnGlu 108
|||||
Db 181 CTCATCCCTCTCAGCACTGATGAAGTTGTAGAGAAGTTGGAGGACATTTTCCAGCAGGAG 240

QY 109 PheSerThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMet 128
|||||
Db 241 TTTCTACACCCCTCAAGGAAGAGTCTGGTACTACAGCTGATCCAGTCTTATCAGCGGATG 300

QY 129 ProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThrMet 148
|||||
Db 301 CCAGGCAACGCTATGTAGGGGCTTCCGGGTATCTATTAAGCGACACGCTGCTCACCATG 360

QY 149 AspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyr 168
|||||
Db 361 GATGACTTGGGTACCTTATATGACAGAACTGGCTCAATGACCAGGTGATGAACATGTAT 420

QY 169 GlyAspLeuValMetAspThrValProGluLysValHisPhePheAsnSerPhePheTyr 188
|||||
Db 421 GGAGATCTGGTCTATGACACAGCTCCCTGAAAGGTGCATTTCTTCAACAGCTTCTTCTAT 480

QY 189 AspLysLeuArgThrLysGlyTyrAspGlyValLysArgTyrThrLysAsnValAspIle 208
|||||
Db 481 GATAAACTCCGTACCAAGGGTTATGATGGGTAAAGAGGTGGACCAAAAATGTGGACATC 540

QY 209 PheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHisTrpSerLeuIleSer 228
|||||
Db 541 TTCAATAAGGAATTACTGGTAATCCCATCCATCTGGAGGTGCACCTGGTCCCTTATCTCA 600

QY 229 ValAspValArgArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgArg 248
|||||
Db 601 GTTGATGTAAAGCGAGCTACCATCATTATTTTGACTCCCGAGGAACTCTTAAATGCCCGC 660

QY 249 CysProLysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAspArgLeuAsp 268
|||||
Db 661 TGCCT-AAGCATATTGC-AAGTATCTACAGGACAGGAGTCAGTCAAAAAGACCGACTGGAC 718

QY 269 PheHisGln 271
|||||
Db 719 TTCCATCAG 727

RESULT 5
BI456217 985 bp mRNA linear EST 21-AUG-2001
LOCUS 603172996F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5252591 5',
DEFINITION mRNA sequence.
ACCESSION BI456217
VERSION BI456217.1 GI:15246873
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 985)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1637 row: o column: 24
High quality sequence stop: 746.
FEATURES
source
Location/Qualifiers
1..985

```

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_image="5252591"  
 /clone\_lib="NCI\_CGAP\_Mam5"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="7 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Robin Humphreys,  
 NIH"

BASE COUNT 263 a 220 c 243 g 258 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 6,62e-136 Length: 985  
 Score: 1133.00 Matches: 208  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.52% Mismatches: 0  
 Query Match: 63.94% Indels: 0  
 DB: 13 Gaps: 0

US-09-848-852A-3 (1-332) x B1456217 (1-985)

QY 124 SerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArg 143

Db 2 TCTATACGGGATGCCAGCACGCTATGGTAAGGGGCTTCGGGTATCCTATAGCGA 61

QY 144 HisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGln 163

Db 62 CACGTGCTCACCATGGTACTTGGTACCTTATATGGACAGAACTGGCTCAATGACCAG 121

QY 164 ValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysValHisPhePhe 183

Db 122 GTGATGAACATGATGAGATCTGGTCATGGACACAGTCCCTGAAAGGTGCATTTCTTC 181

QY 184 AsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTyrPThr 203

Db 182 AACAGCTTCTTCTATGATTAACCTCCGTACCAAGGGTTATGATGGGTAAAGAGTGGACC 241

QY 204 LysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHis 223

Db 242 AAAAATGTGGACATCTTCAATAAGGAATTACTGCTAATCCCATCCATCTGGAGGTGCAC 301

QY 224 TrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAspSerGlnArg 243

Db 302 TGGTCCCTTATCTCACTTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 361

QY 244 ThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGluAlaValLys 263

Db 362 ACTCTAAATCGCGCTGCCCTAAGCATATGCGCAAGTATCTACAGCAGGCGGCTCAAA 421

QY 264 LysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArg 283

Db 422 AAAGACCGACTGGACTTCCATCAGGCTGGAAAGGTTACTTCAAAATGAATGTGGCAGG 481

QY 284 GlnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeu 303

Db 482 CAGAATAATGACAGTCACTGTGGTGGCTTTGTATTACAGTACTGCAAGCACCTGGCCCTG 541

QY 304 SerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgGlnIleTyrLys 323

Db 542 TCTCACCATTACCTTACCCACAGAGACATGCCCCAACTCCGTGTCAGATCTACAAG 601

QY 324 GluLeuCysHisCysLysLeuThrVal 332

Db 602 GAACCTGTGCTCACTGCAAACTCACTGTG 628

RESULT 6  
 BM696661  
 LOCUS 626 bp mRNA linear EST 28-FEB-2002

## DEFINITION

UI-E-DW0-agk-m-22-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone  
 UI-E-DW0-agk-m-22-0-UI 5', mRNA sequence.

## ACCESSION

BM696661

## VERSION

BM696661.1 GI:19009919

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## AUTHORS

1 (bases 1 to 626)

## TITLE

Bonaldo,M.F., Lennon,G. and Soares,M.B.

## JOURNAL

Normalization and subtraction: two approaches to facilitate gene  
 discovery

## MEDLINE

Genome Res. 6 (9), 791-806 (1996)

## COMMENT

97044477

## CONTACT

Contact: Soares, MB

## PROGRAM

Program for Rat Gene Discovery and Mapping

## UNIVERSITY

University of Iowa

## ADDRESS

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

## TEL

Tel: 319 335 8250

## FAX

Fax: 319 335 9565

## EMAIL

Email: msoares@blue.weeg.uiowa.edu

## TISSUE

Tissue Procurement: Dr. Gregg Hageman

## CDNA

CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

## DNA

DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

## CLONE

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

## GENETICS

Clone Distribution: Researchers may obtain clones from Research

## Seq

Genetics (www.resgen.com).

## FEATURES

Seq primer: M13 Reverse.

## LOCATION/QUALIFIERS

1..626

## SOURCE

/organism="Homo sapiens"

## DB\_XREF

/db\_xref="taxon:9606"

## CLONE\_LIB

/clone\_lib="UI-E-DW0-agk-m-22-0-UI"

## TISSUE\_TYPE

/tissue\_type="adult"

## DEV\_STAGE

/dev\_stage="adult"

## LAB\_HOST

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

## NOTE

UI-E-DW0 is a cDNA library containing the following  
 modified polylinker; Site\_1: EcoR I; Site\_2: Not I;

## TISSUE

UI-E-DW0 is a cDNA library containing the following  
 tissue(s): lens. The library was constructed according to

## BONALDO

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. First strand cDNA synthesis was primed with an

## OLIGO-DT

oligo-dT primer containing a Not I site. Double stranded  
 cDNA was ligated to an EcoR I adaptor, digested with Not

## I

I, and cloned directionally into pT73-Pac vector. The  
 oligonucleotide used to prime the synthesis of

## FIRST-STRAND

first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)18 tail. The

## SEQUENCE

sequence tag for this library is CGATTAGCGA. This library  
 was created for the program, Gene Discovery in the Visual

## SYSTEM

System, supported by National Eye Institute (NEI)."

## BASE COUNT

159 a 157 c 172 g 138 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,51e-131 Length: 626  
 Score: 1097.00 Matches: 208  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 61.91% Indels: 0  
 DB: 14 Gaps: 0

US-09-848-852A-3 (1-332) x BM696661 (1-626)

QY 52 GluGluAlaGluArgProGlyGluLysAlaGlyGlnHisSerProLeuArgGluHis 71

Db 2 GAAGAGCAGAGAGGCTGGGAGAAAGCGCCACACAGCCCTCGCAGAGGACAT 61

QY 72 ValThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGlySerLeuPro 91

Db 62 GTGACCTGCTACAGAGCATCTTGGACGAATTCCTTCAAACGTATGGCAGCTCATACC 121



## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1134)  
NIH-MGC <http://mgi.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Lou Staudt  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12273 row: i column: 06  
High quality sequence stop: 583.

## FEATURES

source

Location/Qualifiers

1..1134

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5554637"

/clone\_lib="NIH\_MGC\_85"

/tissue\_type="lymphoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: lymph; Vector: pCMV-SPORT6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 1.867 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC Library."

261 a 326 c 311 g 236 t

## Alignment Scores:

Pred. No.: 1.22e-116 Length: 1134  
Score: 985.50 Matches: 205  
Percent Similarity: 93.27% Conservative: 3  
Best Local Similarity: 91.93% Mismatches: 10  
Query Match: 55.62% Indels: 6  
DB: 14 Gaps: 1

US-09-848-852A-3 (1-332) x BM804530 (1-1134)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAlaHis-LeuAl 20

Db 257 ATGTACTTCGCCAACGGTTTGGGGACATCTGGGCCAGAGGGGCGCAGCTTGGC 316

QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG1 40

Db 317 ACCCCCTGATGCCAGCATCTCATCAGCAATGTGTGCAGCATCGGGGACCATTGTGCCCA 376

QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluY 60

Db 377 GGAGCTTTTTCAGGGCTCAGATTGGGCATGGCAGAGGAGGAGGAGCCCTGGGGAGAA 436

QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80

Db 437 AGCCGCCACAGACGCCCTCGGAGAGGAGCATGTGACCTCGTACAGGATCTTGA 496

QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValAlaGluY 100

Db 497 CGAATTCCTTCAACGATATGGCAGCTCATACCCCTCAGCACTGATGAGGTAGAGAA 556

QY 100 sleuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValAlaG1 120

Db 557 GCTGGAGGACATTTTCCACAGAGATTTCACCCCTTCAGGAAGGGGCTGGTGTGCA 616

QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140

Db 617 GCTGATCCAGTCTTACACGGATGCCAGCAATGCCATGGTGGAGGGGCTTCGAGTGGC 676

QY 140 aTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTriple 160

Db 677 TTATTAGCGCACGCTGCTGACCATGATGACTTGGGACCTTGATGGACAGAACTGGCT 736  
QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180  
Db 737 CAATGACCAAGGTGATGAACATGATGAGACCTTGGTCATGGACATC-CCTGAAAAAGT 795  
QY 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAsp---GlyVa 199  
Db 796 GCATTTCTTCAATAAGTTCTTCTATGATAAATCCGTACCAAGGGTTATGAAGGGGTGA 855  
QY 199 lLysArgTrpThrLysAsnValAspIle-PheAsnLysGlu-LeuLeu-LeuLeuProI 218  
Db 856 AAAAGTGGCACCACAAAAGGGAACCTCTTCAATAAGGAAGTACTGCTGCTAATTCCTCAT 915  
QY 218 e 218  
Db 916 T 916

## RESULT 9

AZ399027/c

LOCUS

DEFINITION

IM0164L10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0164L10 R, DNA sequence.

ACCESSION

AZ399027

VERSION

AZ399027.1 GI:10514099

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 667)

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0164 row: L column: 10

Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 667.

Location/Qualifiers

1..667

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0164L10"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, p-

/note="Vector: PWB42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 161 a 160 c 160 g 186 t  
ORIGIN

## Alignment Scores:

Pred. No.: 3.6e-116 Length: 667  
Score: 979.00 Matches: 188  
Percent Similarity: 92.82% Conservative: 6  
Best Local Similarity: 89.95% Mismatches: 9  
Query Match: 55.25% Indels: 6  
DB: 17 Gaps: 1

US-09-848-852A-3 (1-332) x AZ399027 (1-667)

Qy 126 GlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisVal 145  
Db 665 CAGCGGATGCCAGACACCGCTATGGTAAGGGGCTTCGGGTATCTCTATAAGTGACAGGTG 606  
Qy 146 LeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMet 165  
Db 605 CTCACCATGGATGGTGGTACCTTATATGGACAGAACTGGCTCATGACAGGTGATG 546  
Qy 166 AsnMetTyrGlyAspLeuValMetAspThrValProGluLysValHisPheAsnSer 185  
Db 545 AACATGATGGAGACCTGGTCATGGACACAGTCCTGAAAGGTGCTATCTTCCACAGT 486  
Qy 186 PhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTTrpThrLysAsn 205  
Db 485 CTCCTTCATGATAAACCCTGACCAAGGTTATGATGGGTAAAGAGGTGGACCAAAAT 426  
Qy 206 ValAspPheAsnLysGluLeuLeuLeuPheProIleHisLeuGluValHisTrpSer 225  
Db 425 GTGGACATCTCAATAGGAATTAAGTCTAATCCCATCCATCTGGAAAGTGACGTGTC 366  
Qy 226 LeuIleSerValAspValArgArgThrIleThrTyrPheAspSerGlnArgThrLeu 245  
Db 365 CTTATCTCAGTTGATGTAAGCAACATACCATCACCTATTCTGCTCCAGCAACTTGA 306  
Qy 246 AsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGluAlaVal-LysLysAs 265  
Db 305 AATCGCCGCTGCCCTAAGTATATGTCAGATATCTACAGGACAGGCGAGTCAAAAAAAG 246  
Qy 265 ParGLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAs 285  
Db 245 CCAAGTGGACTTCATCAGGCTGGAAAGGTATTTCAAATGAATGTGGCCAGGCAGAA 186  
Qy 285 nasAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerG1 305  
Db 185 TAATGACAGTCACTGTGGTGCCTTGTGTACAGTACTGCAAGCACCTGGCCCTGTCTCA 126  
Qy 305 nProPheSerPheThrGlnGlnAsp-MetProLysLeuArgArgGlnIleTyrLysGluL 325  
Db 125 GCCATTCAGTTTCACCCAGGAGGACATCT-----CGTCAGATCTGCAAGGAAC 78  
Qy 325 euCysHisCysLysLeuThrVal 332  
Db 77 TCGTGCTACTGCAAACTCACAGTG 55

## RESULT 10

AW916790/c

LOCUS

DEFINITION

EST348198 Rat gene index, normalized rat, norvegicus, Bento Soares

AW916790

ACCESSION

VERSION

AW916790.1 GI:8082631

KEYWORDS

SOURCE

616 bp mRNA linear EST 25-MAY-2000

EST348198 Rat gene index, normalized rat, norvegicus, Bento Soares

AW916790

ACCESSION

VERSION

AW916790.1 GI:8082631

KEYWORDS

SOURCE

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

## REFERENCE

1 (bases 1 to 616)

## AUTHORS

Lee,N.H., Glodok,A.,

Kerlavage,A.R. and Adams,M.D.

Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat

Gene Index

Unpublished (1998)

Contact: Lee, NH

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

This clone is available through the ATCC, contact the ATCC

tel#703-365-2700 for further information

Seq primer: M13 Reverse.

Location/Qualifiers

source

1..616

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

/clone\_lib="Rat gene index, normalized rat, norvegicus,"

Bento Soares"

/tissue\_type="mix - brain, ovary, placenta, kidney, lung,

liver, embryo, heart, muscle, spleen"

/lab\_host="SOLR"

/note="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2:

XhoI; Estimated insert size approx.1 kb"

BASE COUNT 148 a 165 c 153 g 150 t

ORIGIN

## Alignment Scores:

Pred. No.: 1.05e-115 Length: 616  
Score: 975.00 Matches: 187  
Percent Similarity: 95.98% Conservative: 4  
Best Local Similarity: 93.97% Mismatches: 7  
Query Match: 55.02% Indels: 1  
DB: 10 Gaps: 0

US-09-848-852A-3 (1-332) x AW916790 (1-616)

Qy 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAla-HisLeuAl 20  
Db 595 ATGTACTCTACCAATGGTTTGGGGGACTGTCTGTGCTCGAAGGGGAGCGTAGCCTGGC 536  
Qy 20 aProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG1 40  
Db 535 ACCCTTGATGCCAGCATACTCATCAGCAATGTGTGAGCATTTGGACACCGTGGCTCA 476  
Qy 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60  
Db 475 GGAACCTTTTCAGAGCTCGGACTTGGGCACCTGCGAAGAGAGCAGATCGGCAGGGAGAA 416  
Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80  
Db 415 AGTGGCCACATAGACCCCTCGGGGAGAACATGTGACCTGTGTGAGAGTATCTTAGA 356  
Qy 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLy 100  
Db 355 TGAATTCCTCAAACTTATGTAGCTCATCCCTCAGCAGTATGATGAAGTCTGTAGAGAA 296  
Qy 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuG1 120  
Db 295 GTTGGAGGACATTTCCAGCAGGAGTCTCTACCCCTCCAGGAAGAGTCTGTGTACTACA 236  
Qy 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140  
Db 235 GCTGATCCAATCTTACCAGCGGATGCCAGCAACGCATGTAAGGGGCTTCGGGGTATC 176  
Qy 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160



```
Db 175 CTTAAAGACACGCTGCTCACCATGGATGACTGGTACCTTATATGGACAGAACTGGCT 116
      |||
Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
      |||
Db 115 CAATGACGAGGTGATGACATATATGGAGATGTTGTTATGGACACAGTCCCTGAAAGGT 56
      |||
Qy 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGly 198
      |||
Db 55 GCATTTCACACGTTCTCTATGATAAACTCCGTACCAAGGGTTATGATGCC 1
      |||

RESULT 11
AW578424
LOCUS AW578424 620 bp mRNA linear EST 16-MAR-2000
DEFINITION RC1-CT0249-120100-022-c03 CT0249 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW578424
VERSION AW578424.1 GI:7253473
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 620)
JOURNAL HCGP http://www.ludwig.org.br/ORESTES.
COMMENT The FAPESP/LICR Human Cancer Genome Project
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1st2=RC1-CT0249-120100-022-c03&t3=2000-01-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 583.
Location/Qualifiers
1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0249"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 161 a 146 c 163 g 150 t
ORIGIN

Alignment Scores:
Pred. No.: 1.39e-112 Length: 620
Score: 951.00 Matches: 191
Percent Similarity: 91.04% Conservative: 2
Best Local Similarity: 90.09% Mismatches: 2
Query Match: 53.67% Indels: 17
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x AW578424 (1-620)

Qy 61 AlaglyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAsp 80
      |||
Db 12 TCTGGAGACATCG-----GAC 29
      |||
Qy 81 GluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLys 100
      |||
```

```
Db 30 GAATTCCTTAAACGATGTCGACGCTCATATACCCCTCAGCACGTGATGAGTAGAGAA 89
      |||
Qy 101 LeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGln 120
      |||
Db 90 CTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGCCCTGGTGTGCAG 149
      |||
Qy 121 LeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAla 140
      |||
Db 150 CTGATCCAGCTCTACCGCGGATGCCAGCAATGCCATGGTGGGGGCTTCGAGTGGCT 209
      |||
Qy 141 TyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeu 160
      |||
Db 210 TATAAGCGCGCAGCTGCTGACCATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCTC 269
      |||
Qy 161 AsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrVal-ProGluLysVa 180
      |||
Db 270 AATGACGAGGTGATGAACATGATGGAGACCTGGTCATGGACACAGTCCCTGAAAAGGT 329
      |||
Qy 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVal 200
      |||
Db 330 GCATTTCCTCAATAGTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGTGAA 389
      |||
Qy 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeu-IleProIleHisL 220
      |||
Db 390 AAGGTGGACCAAAACGTTACATCTTCAATAAGAGGCTACTGCTAAAATCCCATCCACC 449
      |||
Qy 220 euGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheA 240
      |||
Db 450 TGGAGGTGATGGTCCCTCATCTCTGTTGATGTGAGGCGACGACCATCCTATTATTG 509
      |||
Qy 240 sPserGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyr-LeuGlnAla 259
      |||
Db 510 ACTCGCAGCGTAACCTAAACCGCGCTGCCCTAAGCATATTCGCCAAGTATCTCAGAGCA 569
      |||
Qy 260 GluAlaValLysLysAspArgLeuAspPhe 269
      |||
Db 570 GAGGCGGTAAAGAAAGACCGATGGATTTC 599
      |||

RESULT 12
AW604359 620 bp mRNA linear EST 23-MAR-2000
LOCUS RC1-CT0249-290100-022-c03 CT0249 Homo sapiens cDNA, mRNA sequence.
DEFINITION RC1-CT0249-290100-022-c03 CT0249 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW604359
VERSION AW604359.1 GI:7309100
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS 1 (bases 1 to 620)
TITLE HCGP http://www.ludwig.org.br/ORESTES.
JOURNAL The FAPESP/LICR Human Cancer Genome Project
COMMENT Unpublished (1999)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1st2=RC1-CT0249-290100-022-c03&t3=2000-01-29&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 583.
Location/Qualifiers
1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0249"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 161 a 146 c 163 g 150 t
ORIGIN

Alignment Scores:
Pred. No.: 1.39e-112 Length: 620
Score: 951.00 Matches: 191
Percent Similarity: 91.04% Conservative: 2
Best Local Similarity: 90.09% Mismatches: 2
Query Match: 53.67% Indels: 17
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x AW578424 (1-620)

Qy 61 AlaglyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAsp 80
      |||
Db 12 TCTGGAGACATCG-----GAC 29
      |||
Qy 81 GluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLys 100
      |||
```

/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 161 a 146 c 163 g 150 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,39e-112 Length: 620  
Score: 951.00 Matches: 191  
Percent Similarity: 91.04% Conservative: 2  
Best Local Similarity: 90.09% Mismatches: 2  
Query Match: 53.67% Indels: 17  
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x AW604359 (1-620)

QY 61 AlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAsp 80  
Db 12 TCAGAGGACATTGG-----GAC 29  
QY 81 GluPheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluValValGluLys 100  
Db 30 GAATTCCTTCAACGATATGCGACCTCATACCCCTCAGCACTCATGGGTAGAGAAG 89  
QY 101 LeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGln 120  
Db 90 CTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGCCCTGGTTCGAG 149  
QY 121 LeftLegInSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheArgValAla 140  
Db 150 CTGATCAGTCTTACCAGCGGATGCCAGGCAATGCCATGTGTGAGGGCTTCCAGTGGCT 209  
QY 141 TyrlsArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrglnAsnTrpLeu 160  
Db 210 TATAAGCGGCACGTGCTGACCATGGATGACTTGGGACCTTGTATGGACAGAAGTGGCTC 269  
QY 161 AsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThrVal-ProGluLysVa 180  
Db 270 AATGACGAGTGATGACATGTATGGAGAGCTGTGATGGACACAGTCCCTGAAAGGT 329  
QY 180 lHisPheAsnSerPhePheTyrglyAspLysLeuArgThrLysGlyTyrglyVally 200  
Db 330 GCATTTCTTCAATAGTTTCTTATGATAAATCCCGTACCAAGGGTATGATGGGTGAA 389  
QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeu-IleProIleHisL 220  
Db 390 AAGGTGGACCAAAACGTGTACATCTTCAATAAGGAGCTACTGCTAAATCCCATCCACC 449  
QY 220 euGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrglyPheA 240  
Db 450 TGGAGGTGATTTGGTCCCTCATCTCTGTTGATGTGAGGGCAGCACCACCATATTTTG 509  
QY 240 spSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrglyGlnAla 259  
Db 510 ACTCGAGCGTAACCTAAACCGCGCTGCCCTAAGCATATTTGCCAAGTATTCCTACAGGCA 569  
QY 260 GluAlaValLysLysAspArgLeuAspPhe 269  
Db 570 GAGGCGTAAAGAAAGACCGGATTGGATTTC 599

RESULT 13  
BE077141  
LOCUS BE077141 572 bp mRNA linear EST 09-JUN-2000  
DEFINITION RC5-BT0604-150200-031-E07 BT0604 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE077141  
VERSION BE077141.1 GI:8427661  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 572)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., F.,D.H., Goldmann,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC5-BT0604-150 200-031-E07&ts=2000-02-15&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 499.

FEATURES  
source  
1..572  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BT0604"  
/dev\_stage="Adult"  
/note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 140 a 143 c 149 g 140 t

ORIGIN  
Alignment Scores:  
Pred. No.: 8,79e-110 Length: 572  
Score: 929.00 Matches: 175  
Percent Similarity: 88.83% Conservative: 0  
Best Local Similarity: 88.83% Mismatches: 0  
Query Match: 52.43% Indels: 22  
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x BE077141 (1-572)

QY 136 GlyPheArgValAlaTyrlsArgHisValLeuThrMetAspLeuGlyThrLeuTy 155  
Db 26 GCGTTTCGAGTTGCTTATAAGCGGCACGTGCTGACCATGGATGACTGGGACCTTGAT 85  
QY 156 GlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThr 175  
Db 86 GGACAGAACTGGCTCAATGACGAGGTGATGAACATGTATGGAGACCTGTGTCATGGACACA 145  
QY 176 ValProGluLysValHisPhePheAsnSerPhePheTyrglyAspLysLeuArgThrLysGly 195  
Db 146 GTCCCTGAAAGAGTGCATTTCTTCAATAGTTTC----- 178  
QY 196 TyrAspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeu 215  
Db 179 -----TTCAATAAGGAGCTACTGCTA 199  
QY 216 IleProIleHisLeuGluValHisTrpSerLeuIleSerValAspValArgArgThr 235  
Db 200 ATCCCCATCCACCTGGAGGTGATGGTCCCTCATCTCTGTTGATGTGAGGCGACGCACC 259

Qy	236	IleThrTyrPheAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLys	255
Db	260	ATCACCTATTATTGACTCGCAGCGTACCCTAAACCGCGCTGCCCTAAGCATATTGCCAAG	319
Qy	256	TyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrrLysGly	275
Db	320	TATCTACAGCAGAGGCGGTAAAGAAAGACCGACTGGATTCTCCACCAAGGCTGGAAGGT	379
Qy	276	TyrPheLysMetAsnValAlaArgGlnAsnAspSerAspCysGlyAlaPheValLeu	295
Db	380	TACTTCAAAATGATGTGGCCAGCAGAAATGACAGTACTGTGGTGCTTTTGTGTG	439
Qy	296	GlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetPro	315
Db	440	CAGTACTGCAAGCATCTGGCCCTCTCTCAGCATTCAGTTCACCCAGGACATGCC	499
Qy	316	LysLeuArgGlnIleTyrLysGluLeuCysHisCysLysLeuThrVal	332
Db	500	AAACTTCGTGGCAGATCTACAAGGAGCTGTCTACTGCAAACTCACTGTG	550
RESULT	14		
AL578220/c			
LOCUS	AL578220	626 bp mRNA linear	EST 16-FEB-2001
DEFINITION	AL578220 LTI_NFL006_PL2	Homo sapiens cDNA clone CS0DK002YD13	3
		prime, mRNA sequence.	
ACCESSION	AL578220		
VERSION	AL578220.1	GI:12942089	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
FEATURES	Location/Qualifiers		
source	1..626		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CS0DK002YD13"		
	/clone_1lb="LTI_NFL006_PL2"		
	/tissue_type="placenta"		
	/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"		
BASE COUNT	142 a 143 c 165 g 164 t	12 others	
ORIGIN			
Alignment Scores:			
Pred. No.:	4.54e-109	Length:	626
Score:	924.00	Matches:	178
Percent Similarity:	90.86%	Conservative:	1
Best Local Similarity:	90.36%	Mismatches:	18
Query Match:	52.14%	Indels:	1
DB:	9	Gaps:	0
US-09-848-852A-3	(1-332) x AL578220	(1-626)	
Qy	136	GlyPheArgValAlaTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyr	155
Db			

Stem cell origin."  
 /lab\_host="DH10B"  
 /notes="Organ: lung; Vector: PCMV-SPORT6; Site.1: SalI;  
 Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

BASE COUNT 232 a 247 c 190 g 201 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.08e-107 Length: 870  
 Score: 913.00 Matches: 190  
 Percent Similarity: 88.79% Conservative: 8  
 Best Local Similarity: 85.20% Mismatches: 13  
 Query Match: 51.52% Indels: 12  
 DB: 13 Gaps: 3

US-09-848-852A-3 (1-332) x BII52305 (1-870)

```

Qy 121 LeuileGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAla 140
Db 1 CTGATCCAGTCTTATCAGCGGATGCCAGGCAACGCTATGGTAAGGGGCTTCCGGGTATCC 60

Qy 141 TyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeu 160
Db 61 TATAAGCGACACGCTGCTCACCATGGATGACTTGGGTACCTTATATGGACAGAACTGGCTC 120

Qy 161 AsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVal 180
Db 121 AATGACCAGGTGATGAACATGTATGGAGATCTGGTCATGGACACAGTCCCTGAAAGGGT 180

Qy 181 HisPhe-PheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
Db 181 CATTATCTTCAACAGCTTCTTCATGATATAACTCCGTACCAAGGTTATGATGGGGTAAA 240

Qy 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220
Db 241 GAGGTGGACCAAAAATGTGGACATCTTCAATAAGGAATTACTGCTAATCCCATCCATCT 300

Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240
Db 301 GGAGTGCACTGGTCCCTTATCTCAGTTGATGTAAGGCGACGTACCATCACCCTATTTTGA 360

Qy 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaG1 260
Db 361 CTCCACGGAACCTAAATCGCGCTGCCCTAAGCATATTGCCAAGTATCTACACGCCCA 420

Qy 260 uAlaValLysLysAspArgLeuAspPhe-----HisGln-GlyTrp-LysGlyTyrPhe 277
Db 421 GACGGCCAGTCAAAACACGACCGACTGGACTATCCATCAGCGGTGGACAAGGTACTTC 480

Qy 278 LysMetAsnValAla-ArgGlnAsn---AsnAspSerAspCys-GlyAla---PheValL 295
Db 481 ACATGAATGTGGCCAGCGGACAGACATACACTGACATGACTGCTGGATGCCCTTATGTAT 540

Qy 295 euGlnTyrCys-LysHisLeuAlaLeu-SerGlnPro-PheSerPheThrGlnGlnAspM 314
Db 541 CACAGTACTGCCAAGCACCTGGCCCTGATCTCAGCCACTTCAGCTTCACCCACGAGGACA 600

Qy 314 etProLysLeuArgArgGlnIleTyrLysGluLeuCysHisCysLysLeuThr 331
Db 601 TGCCCAAACTCCGTCGTGATCTACAAGGAAGTGTGTCTACTGCAAACTCACT 653

```

Search completed: December 16, 2002, 23:01:05  
 Job time : 2428 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2002, 14:18:34 ; Search time 397 Seconds  
(without alignments)  
11294.027 Million cell updates/sec

Title: US-09-848-852A-4  
Perfect score: 1991  
Sequence: 1 GCCCTCCCTGTCCCGACCC.....TCCTCTGCTCAATAAGATC 1991

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1925.4	96.7	2080	22	AAI58425 Human polynucleoti
2	1843.4	92.6	2206	22	AAI58431 Nucleotide sequenc
3	1813.4	91.1	2029	21	AAI58426 Human polynucleoti
4	1667	83.7	1760	21	AAZ43782 Human fetal brain
5	1568.2	78.8	1804	22	AAI60211 Human polynucleoti
6	1568.2	78.8	1804	22	AAI60212 Human polynucleoti
7	1451.2	72.9	1934	22	AAI60213 Human secreted pro
8	1440.8	72.4	1933	22	AAI60214 Human ovarian anti
9	1357.4	68.2	1507	22	AAI60215 Human secreted pro

10	567.4	28.5	650	21	AAC59569 Human secreted pro
11	551	27.7	695	22	AAD05540 Human secreted pro
12	486.4	24.4	5318	22	AAF55961 Human eukaryotic i
13	468.4	23.5	870	22	AAI94288 Human neuroblastom
14	442	22.2	503	22	AAI25529 Human ovarian PCR-
15	362.2	18.2	820	22	AAI93958 Human neuroblastom
16	344.6	17.3	714	22	AAH03368 Human cDNA clone (
17	344.6	17.3	1578	22	AAH13708 Human CDNA sequenc
18	291.2	14.6	358	22	AAI84371 Human polynucleoti
19	229.6	11.5	1938	22	AAH99558 Human protein enco
20	208.2	10.5	529	22	AAD05542 Human secreted pro
21	181	9.1	181	21	AAC08625 Human secreted pro
22	177	8.9	208	21	AAA45202 Human secreted exp
23	167.4	8.4	855	22	AAI94287 Human neuroblastom
24	115.8	5.8	501	22	AAI61344 Human immune/haema
25	98.4	4.9	2512	22	AAF54827 Nucleotide sequenc
26	92	4.6	435	24	ABL81757 Human ovarian canc
27	91.8	4.6	666	23	ABSL71205 DNA encoding novel
28	86	4.3	833	22	AAI93957 Human neuroblastom
29	84.8	4.3	106	24	ABL81741 Human ovarian canc
30	72	3.6	14551	24	ABL34585 Human metastasis a
31	70.8	3.6	11745	24	ABK28332 DNA transcription
32	70.6	3.5	147	22	AAI61721 Soybean 515002 reg
33	70.6	3.5	150	22	Soybean 240017 reg
34	70.6	3.5	127197	22	Soybean 515002 reg
35	70.6	3.5	335913	22	AAI61371 Soybean 240017 reg
36	70.6	3.5	335913	22	AAI61371 Soybean 240017 reg
37	70.6	3.5	335913	22	AAI61372 Soybean 240017 reg
38	70.6	3.5	335913	22	AAI61372 Soybean 240017 reg
39	70.4	3.5	284	22	AAI61568 Soybean 318013 reg
40	70.4	3.5	513445	22	AAI61373 Soybean 318013 reg
41	70.2	3.5	274	19	AAV16132 Microsatellite mar
42	70	3.5	51952	19	AAV26084 Tomato pest resist
43	69.8	3.5	6298	22	AAF75767 Promoter of gene r
44	69.6	3.5	4049	23	ABL50553 Mial2 nucleotide s
45	69.6	3.5	8952	22	RAA46446 Tumour suppressor

## ALIGNMENTS

## RESULT 1

AAI58425  
ID AAI58425 standard; cDNA; 2080 BP.

XX AAI58425;

XX 22-OCT-2001 (first entry)

DT Human polynucleotide SEQ ID NO 628.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.  
(HYSE-) HYSEQ INC.  
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
WPI: 2001-442253/47.  
P-PSDB; AAM39269.  
Novel nucleic acids and polypeptides, useful for treating disorders  
such as central nervous system injuries -  
Claim 1; SEQ ID NO 628; 10078pp; English.  
The invention relates to human nucleic acids (AA157798-AA161369) and  
the encoded polypeptides (AAM38642-AA442213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide of polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 2080 BP; 472 A; 567 C; 570 G; 471 T; 0 other;  
Query Match 96.7%; Score 1925.4; DB 22; Length 2080;  
Best Local Similarity 98.9%; Pred. No. 0;  
Matches 1967; Conservative 0; Mismatches 1; Indels 21; Gaps 2;  
QY 1 GCCTCCCTGTCGCCGACCTCTTTTCATGCTCAGCAAGTGAAGAGGAGGAGGAGG 60  
DB 93 GCCTCCCTGTCGCCGACCTCTTTTCATGCTCAGCAAGTGAAGAGGAGGAGGAGG 152  
QY 61 AGGAGGAGGAGGAGTGAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120  
DB 153 AGGAGGAGGAGGAGTGAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 212  
QY 121 GGAGTCAGCTGGGAACCTCCGAGGCGCCGCGCTTCCGCGCCGACATCATGCAAAACCT 180  
DB 213 GGAGTCAGCTGGGAACCTCCGAGGCGCCGCGCTTCCGCGCCGACATCATGCAAAACCT 272  
QY 181 GCTCAGAGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240  
DB 273 GCTCAGAGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 332  
QY 241 CCTCGCTGACATTCACATGGAAGCTTTGGGGGCGCCACCGGGGCGCGGGGCGCTCG 300  
DB 333 CCTCGCTGACATTCACATGGAAGCTTTGGGGGCGCCACCGGGGCGCGGGGCGCTCG 392  
QY 301 CACACCCCAAGAACCATCTTTTCAACCCAGCAAGGGGAGGAGGAGGAGGAGGAGGAGG 360  
DB 393 CACACCCCAAGAACCATCTTTTCAACCCAGCAAGGGGAGGAGGAGGAGGAGGAGGAG 452  
QY 361 CCTGTTCTGCTTTTGTACCTCCCGGGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 420  
DB 453 CCTGTTCTGCTTTTGTACCTCCCGGGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 512  
QY 421 CTCTCATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480  
DB 513 CTCTCATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 572  
QY 481 AGGAAGATGACATGAGTGGACATCCAAAGTCTCCTCTGAGACCTGACTCGGGGCTCCTTT 540  
DB 1712

DB 573 AGGAAGATGAGTACTAGGTGGACTCCAAAGATCTCCTCTGGACCTTGACTCGGGGCTCCTTT 632  
QY 541 CATGTACTCTGCCCAACGGTTTTGGGGGACAATCTGGGGCAGAAAGGAGGAGCGCA-CTTGG 599  
DB 633 CATGTACTCTGCCCAACGGTTTTGGGGGACAATCTGGGGCAGAAAGGAGGAGCGCAGCTTG 692  
QY 600 CACCCCTGATGCGCAGCATCCTCATCAGCAATGTGTGAGCATCGGGGAGGAGCATGTGCCCC 659  
DB 693 CACCCCTGATGCGCAGCATCCTCATCAGCAATGTGTGAGCATCGGGGAGGAGCATGTGCCCC 752  
QY 660 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGGAGGAGGAGGAGGAGGAG 719  
DB 753 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGGAGGAGGAGGAGGAGGAG 812  
QY 720 AAGCGGCCAGCAGACGCCCCCTGCGAGGAGGAGCATGTGACCTGCGTACAGAGCATCTTGG 779  
DB 813 AAGCGGCCAGCAGACGCCCCCTGCGAGGAGGAGCATGTGACCTGCGTACAGAGCATCTTGG 872  
QY 780 ACGAATTCCTTCAAAACGATATGGCAGCCTCATACCCCTCAGCACTGTATGAGTATAGAGA 839  
DB 873 ACGAATTCCTTCAAAACGATATGGCAGCCTCATACCCCTCAGCACTGTATGAGTATAGAGA 932  
QY 840 AGCTGAGGAGCATTTTCCAGCAGAGATTTTCCACCCCTTCCAGAGAGGAGGAGGAGGAG 899  
DB 933 AGCTGAGGAGCATTTTCCAGCAGAGATTTTCCACCCCTTCCAGAGAGGAGGAGGAGGAG 992  
QY 900 AGCTGATCCAGTCTTACACGCGATGCGAGCAATGCGATGGTGAGGAGGAGGAGGAGGAG 959  
DB 993 AGCTGATCCAGTCTTACACGCGATGCGAGCAATGCGATGGTGAGGAGGAGGAGGAGGAG 1052  
QY 960 CTTATAGCGCGCAGCTGCTGACCATGGATGACTTGGGAGCCTTGTATGAGACAGAACTGGC 1019  
DB 1053 CTTATAGCGCGCAGCTGCTGACCATGGATGACTTGGGAGCCTTGTATGAGACAGAACTGGC 1112  
QY 1020 TCAATGACGAGGTGATGAACATGTATGGAGACCTGCTGATGGACACAGTCCCTGAAAAGG 1079  
DB 1113 TCAATGACGAGGTGATGAACATGTATGGAGACCTGCTGATGGACACAGTCCCTGAAAAGG 1172  
QY 1080 TGCAATTTCTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGGTGA 1139  
DB 1173 TGCAATTTCTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGGTGA 1232  
QY 1140 AAAGGTGGACCAAAACGAGTGGACATCTTCAATAGGAGCTACTGCTAATCCCATCCAC 1199  
DB 1233 AAAGGTGGACCAAAACGAGTGGACATCTTCAATAGGAGCTACTGCTAATCCCATCCAC 1292  
QY 1200 TGAGGTGCAATTTGGTCCCTCATCTCTGTTGATGTGAGGCGACGACCATCACTATTTTG 1259  
DB 1293 TGGAGGTGCAATTTGGTCCCTCATCTCTGTTGATGTGAGGCGACGACCATCACTATTTTG 1352  
QY 1260 ACTCGAGCGCTACCCCTAAACCGCGCTGCGCTTAAGCATATTGCCAAGTATCTACAGGCG 1319  
DB 1353 ACTCGAGCGCTACCCCTAAACCGCGCTGCGCTTAAGCATATTGCCAAGTATCTACAGGCG 1412  
QY 1320 AGCGGTAAAGAAAGACCGGACTGGATTTCCACGAGGCTGGAAAGGTTACTTCAAAATGA 1379  
DB 1413 AGCGGTAAAGAAAGACCGGACTGGATTTCCACGAGGCTGGAAAGGTTACTTCAAAATGA 1472  
QY 1380 ATGTGGCCAGGAGAGATAATGACAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1439  
DB 1473 ATGTGGCCAGGAGAGATAATGACAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1532  
QY 1440 ATCTGCGCTGCTCAGCCATTGAGCTTCCAGGAGGAGCATGCCCAAACTTCGTCGGC 1499  
DB 1533 ATCTGCGCTGCTCAGCCATTGAGCTTCCAGGAGGAGCATGCCCAAACTTCGTCGGC 1592  
QY 1500 AGATCTACAGAGGAGTGTGTCTACTGCAAACTCACTGTGTGAGGCTGTGTGTGTGTGTGTGT 1559  
DB 1593 AGATCTACAGAGGAGTGTGTCTACTGCAAACTCACTGTGTGAGGCTGTGTGTGTGTGTGTGT 1652  
QY 1560 AAGCCCATAAATGGGAGGAGAGCATGGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC 1619  
DB 1653 AAGCCCATAAATGGGAGGAGAGCATGGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC 1712

```
QY 1620 CTCCTTGCTCTCCCACTCACTCCCTTGGTTTTCATATTTAAATGTTCAATTC 1679
|||||
Db 1713 CTCCTTGCTCTCCCACTCACTCCCTTGGTTTTCATATTTAAATGTTCAATTC 1772
QY 1680 TGTATTTTTCCTTTGAGAGAACTACTGTTGATTTCTGATGTCAGGGGTGCTACA 1739
|||||
Db 1773 TGTATTTTTCCTTTGAGAGAACTACTGTTGATTTCTGATGTCAGGGGTGCTACA 1832
QY 1740 GAAAGCCCTTCTCTCTCTGTTGAGGGAGTGGCCCTGTCCTGGTGGAGCA 1799
|||||
Db 1833 GAAAGCCCTTCTCTCTCTGTTGAGGGAGTGGCCCTGTCCTGGTGGAGCA 1892
QY 1800 GTCATCCCTCCCTTCCCGTGCAGGGAGCAGGAATCAGTCTCGGGGTGTCGGCGGA 1859
|||||
Db 1893 GTCATCCCTCCCTTCCCGTGCAGGGAGCAGGAATCAGTCTCGGGGTGTCGGCGGA 1952
QY 1860 CAATAGGATCACTGCTGCGAGATCTCAAACTTTATATATATATATATATATATAT 1919
|||||
Db 1953 CAATAGGATCACTGCTGCGAGATCTCAAACTTTT-----TATAT 1992
QY 1920 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1979
|||||
Db 1993 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2052
QY 1980 TCAATAAAG 1988
|||||
Db 2053 TCAATAAAG 2061

RESULT 2
AAF54831
ID AAF54831 standard; DNA; 2206 BP.
XX AC AAF54831;
XX DT 15-MAY-2001 (first entry)
XX DE Nucleotide sequence of a human SENP2 polypeptide.
XX KW Human; de-sentrinase; SENP; SENP1; SENP2; SENP3; chromosome 12q13.1;
KW sentrin-specific protease; sentrin; anti-proliferative agent;
KW anti-viral agent; herpes simplex-1; cytomegalovirus; sentrinisation;
KW PML; tumour suppressor; acute promyelocytic leukaemia; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 52..1758
FT FT /*tag= a
FT FT /product= "SENP3"
XX PN WO200109292-A2.
XX PD 08-FEB-2001.
XX PF 31-JUL-2000; 2000WO-US20884.
XX PR 31-JUL-1999; 99US-0146774.
XX PA (TEXA ) UNIV TEXAS SYSTEM.
XX PI Gong L, Yeh ETH;
XX DR WPI; 2001-182947/18.
XX DR P-PSDB; AAB31977.
XX PT New polynucleotide encoding de-sentrinase polypeptides, useful as
XX PT antiproliferative or antiviral agents -
XX PS Claim 36; Page 120; 122pp; English.
XX CC The present sequence encodes a human de-sentrinase (SENP) 3 polypeptide.
```

CC The specification describes SENP1 and SENP2 polypeptides. The SENP1  
CC gene is found in chromosome 12q13.1. SENP1 is found in the nucleus,  
CC SENP3 is found in the nucleolus, and SENP1 is found in both locations.  
CC SENP polypeptides are sentrin-specific proteases that remove sentrin  
CC from some sentrinised peptides, but do not affect proteins modified by  
CC ubiquitin or NEDD8. SENP polypeptides are used to identify specific  
CC modulators of SENP. These modulators are potential anti-proliferative  
CC and anti-viral agents such as herpes simplex-1 or cytomegalovirus.  
CC SENP polypeptides are used for studying the role of sentrinisation in  
CC the biological function of PML, a tumour suppressor implicated in  
CC development of acute promyelocytic leukaemia. Fragments of SENP  
CC polynucleotides are used as hybridisation probes and amplification  
CC primers for detecting gene expression or preparing mutated sequences,  
CC also as antisense sequences for inhibiting SENP expression.

SQ Sequence 2206 BP; 491 A; 612 C; 605 G; 498 T; 0 other;

Query Match 92.6%; Score 1843.4; DB 22; Length 2206;  
Best Local Similarity 96.8%; Pred. No. 0;  
Matches 1925; Conservative 0; Mismatches 31; Indels 33; Gaps 3;

```
QY 1 GCCTCCCTGTCCTCCCGACCTCTCTTTGATGCTCAGCAAGTGAAGAGGAGGAGAGG 60
|||||
Db 230 GCCTCCCTGTCCTCCCGACCTCTCTTTGATGCTCAGCTAGTGAAGAGGAGGAGAGG 289
QY 61 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGTGGAGGCTTGGAGGCTGCCCAAGAT 120
|||||
Db 290 AGGAG-----GAAGATGAGGAGGAGTGGAGGCTTGGAGGCTTACCCCTAGGT 337
QY 121 GGAGTCACTGGGAACCTCCAGCGGCCCTTCCCGCCCTCATCATGCAAAACCT 180
|||||
Db 338 GGGGCAACTGGGGGCTCCAGCGCTCTCGAGCTCTCCGACCTCTCATGAAAAACCT 397
QY 181 GCTACAGCGCGCGCGAGCCATGAGAGCTTCCGGATGCTGCTCTACTCAAAAGCA 240
|||||
Db 398 GCTACAGCGCGCGCGAGCCATGAGAGCTTCCAGATGCTGCTCTACTCAAAAGCA 457
QY 241 CCTCGCTGACATTCACCTGGAACTTTGGGGGCGCCACCGGGCGCGGGCGGCGCTCG 300
|||||
Db 458 CCTCGCTGACATTCACCTGGAACTTTGGGGGCGCCACCGGGCGCGGGCGGCGCTCG 517
QY 301 CACACCCCAAGAACCATCTTTACCCCAAGAGGGGTGCGACGCCACAGGTGCCATCCC 360
Db 518 CACACCCCAAGAACCATCTTTACCCCAAGAGGGGTGCGACGCCACAGGTGCCATCCC 577
QY 361 CCTGTTGCTGTTTACCTCCCGCGGCGCCACCTCCACCGCGCTGGGTCTGTAGTG 420
|||||
Db 578 CCTGTTGCTGTTTACCTCCCGCGGCGCCACCTCCACCGCGCTGGGTCTGTAGTG 637
QY 421 CTCTCATGGCTGAGGATGGGTGAGAGGCTCTCCACCACTGCTGCGGCCCTCCATGG 480
|||||
Db 638 CTCTCATGGCTGAGGATGGGTGAGAGGCTCTCCACCACTGCTGCGGCCCTCCATGG 697
QY 481 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTGGACCTGACTGGGGCTCTCTTT 540
|||||
Db 698 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTGGACCTGACTGGGGCTCTCTTT 757
QY 541 CATGTACTCTGCCCAACGGTTTGGGGGACATCTGGGCCAGAGAGGGAGCGCA-CTTGG 599
|||||
Db 758 CATGTACTCTGCCCAACGGTTTGGGGGACATCTGGGCCAGAGAGGGAGCGCACTTGG 817
QY 600 CACCCCTGATGCCAGCATCTCTCATCAGCAATGTGTGACGATCGGGGACCATGTGGGCC 659
|||||
Db 818 CACCCCTGATGCCAGCATCTCTCATCAGCAATGTGTGACGATCGGGGACCATGTGGGCC 877
QY 660 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGAGGAGGCGCTGGGGAGA 719
Db 878 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGAGGAGGCGCTGGGGAGA 937
QY 720 AAGCGCGCAGCAGAGCCCTTGGCGAGGAGCATGTGACCTGCGTACAGACATCTTGG 779
|||||
Db 938 AAGCGCGCAGCAGAGCCCTTGGCGAGGAGCATGTGACCTGCGTACAGACATCTTGG 997
```

Qy	780	ACGAATTCCTTTCAAAGTGATGGCAGCGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA	839
Db	998	ACGAAATTCCTTTCAAAGTGATGGCAGCGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA	1057
Qy	840	AGCTGGAGGACATTTTCCACGACGAGGCTTTTCCACCCTTCCAGGAAGGCGCTGGTGTGC	899
Db	1058	AGCTGGAGGACATTTTCCACGACGAGGCTTTTCCACCCTTCCAGGAAGGCGCTGGTGTGC	1117
Qy	900	AGCTGATCCAGTCTTACCACGGGATGCCAGGCAATGCCATGGTGAAGGGCTTCCGAGTGG	959
Db	1118	AGCTGATCCAGTCTTACCACGGGATGCCAGGCAATGCCATGGTGAAGGGCTTCCGAGTGG	1177
Qy	960	CTTATAAGCGGCACGTCTGACCATGGATGACTTGGGACCTTGTATGGACAGAACTGGC	1019
Db	1178	CTTATAAGCGGCACGTCTGACCATGGATGACTTGGGACCTTGTATGGACAGAACTGGC	1237
Qy	1020	TCAATGACCAGGTGATGAACATGTATGGAGAACCTGGTCATGGACACAGTCCCTGAAAGG	1079
Db	1238	TCAATGACCAGGTGATGAACATGTATGGAGAACCTGGTCATGGACACAGTCCCTGAAAGG	1297
Qy	1080	TGCATTTCTTCAATAGTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGTGA	1139
Db	1298	TGCATTTCTTCAATAGTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGTGA	1357
Qy	1140	AAAGGTGGACAAAAACGTGGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACC	1199
Db	1358	AAAGGTGGACAAAAACGTGGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACC	1417
Qy	1200	TGAGGTGCATTTGGTCCCTCATCTCTGTGTGATGTGAGGGGACGACATCACTATTTTG	1259
Db	1418	TGAGGTGCATTTGGTCCCTCATCTCTGTGTGATGTGAGGGGACGACATCACTATTTTG	1477
Qy	1260	ACTCGACGCTACCTTAAACCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAGCGAG	1319
Db	1478	ACTCGACGCTACCTTAAACCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAGCGAG	1537
Qy	1320	AGCGGTAAAGAAAGACCGACTGGATTTCCACCAGGGCTGGAAAGTTTACTTCAAAATGA	1379
Db	1538	AGCGGTAAAGAAAGACCGACTGGATTTCCACCAGGGCTGGAAAGTTTACTTCAAAATGA	1597
Qy	1380	ATGTGCCCAGGCAGAAATATGACAGTGACTGTGGTGCTTTTGTGTGCAAGTACTGCAAGC	1439
Db	1598	ATGTGCCCAGGCAGAAATATGACAGTGACTGTGGTGCTTTTGTGTGCAAGTACTGCAAGC	1657
Qy	1440	ATCTGGCCCTGTCTCAGCCATTTCAGCTTCACCCACGAGACATGCCAAACTTCGTGGCC	1499
Db	1658	ATCTGGCCCTGTCTCAGCCATTTCAGCTTCACCCACGAGACATGCCAAACTTCGTGGCC	1717
Qy	1500	AGATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTGTGAGCCCTCGTACCCAGACCCC	1559
Db	1718	AGATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTGTGAGCCCTCGTACCCAGACCCC	1777
Qy	1560	AAGCCCATAAATGGGAAGGGAGACATGGGAGTCCCTTCCCAAGAAACTTCAGTTCCTTTC	1619
Db	1778	AAGCCCATAAATGGGAAGGGAGACATGGGAGTCCCTTCCCAAGAAACTTCAGTTCCTTTC	1837
Qy	1620	CTCTCTTGCCCTCTTCCCATCTCACTTCCCTTTCGGTTTTTTCATATTAAATGTTTCAATTC	1679
Db	1838	CTCTCTTGCCCTCTTCCCATCTCACTTCCCTTTCGGTTTTTTCATATTAAATGTTTCAATTC	1897
Qy	1680	TGTAATTTTTTTTTTCTTTTGAGAAATACTTGTGATTTCTGATGTGCAAGGGGTGGCTACA	1739
Db	1898	TGTAATTTTTTTTTTCTTTTGAGAAATACTTGTGATTTCTGATGTGCAAGGGGTGGCTACA	1957
Qy	1740	GAAAAGCCCTTCTTCTCTCTTTTGACAGGGAGTGTGGCCCTGTGGCCTGGGTGGAGCA	1799
Db	1958	GAAAAGCCCTTCTTCTCTCTTTTGACAGGGAGTGTGGCCCTGTGGCCTGGGTGGAGCA	2017
Qy	1800	GTCACTCTCCCTCTCCCGTGCAGGGAGCAGGAATCAGTCTGGGGGTGTGGCGGGA	1859
Db	2018	GTCACTCTCCCTCTCCCGTGCAGGGAGCAGGAATCAGTCTGGGGGTGTGGCGGGA	2077
Qy	1860	CAATAGGATCACTGCCTGCCAGATCTTCAAACTTTTATATATATATATATATATATAT	1919

Db	2078	CAATAGATCACTGCCTGCCAGATCTTCAAACCTTT-----TATAT	
Qy	1920	ATATATATATATATATATATATATATAAATGACCACGGTCCTGCTCTGG	
Dd	2118	ATATATATATATATATATATATATATAAAATATATAAATGCCACGGTCCTGCTCTGG	
Qy	1980	TCAATAAAG 1988	
Dd	2178	TCAATAAAG 2186	
<b>RESULT 3</b>			
ID	AAI58426	standard; cDNA; 2029 BP.	
XX	AAI58426;		
XX	22-OCT-2001	(first entry)	
XX	Human polynucleotide SEQ ID NO 629.		
XX	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia; ss.		
OS	Homo sapiens.		
XX	WO200153312-A1.		
XX	26-JUL-2001.		
XX	26-DEC-2000; 2000WO-US34263.		
XX	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
XX	29-NOV-2000; 2000US-0727344.		
FA	(HYSE-) HYSEQ INC.		
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX	WPI; 2001-442253/47.		
DR	P-PSDB; AAM39270.		
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
FS	Claim 1; SEQ ID NO 629; 10078pp; English.		
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and		
CC	the encoded polypeptides (AAM38642-AAM42213), with neurotropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are used		
CC	in gene therapy. A composition containing a polypeptide or polynucleo-		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	activin/inhibin activity, chemotactic/chemokinetic activity, haemostat-		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening		
CC	assays for receptor activity, arthritis and inflammation, leukaemias ac		



CC C.N.S disorders.  
CC Note: the sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 2029 BP; 460 A; 560 C; 553 G; 456 T; 0 other;

Query Match 91.1%; Score 1813.4; DB 22; Length 2029;  
Best Local Similarity 96.3%; Pred. No. 0;  
Matches 1916; Conservative 0; Mismatches 1; Indels 72; Gaps 3;

QY	1	GCCTCCTCTCCCCGACCCCTCTTTTGATGCTCCTCAGCAAGTGAAGAGGAGGAAGAGG	60
DB	93	GCCTCCTCTCCCCGACCCCTCTTTTGATGCTCCTCAGCAAGTGAAGAGGAGGAAGAGG	152
QY	61	AGGAGGAGGAGTGAAGTGAAGAGGAGGAGTGGACCTTGGAGGCTGCCCCCAAGAT	120
DB	153	AGGAGGAGGAGTGAAGTGAAGAGGAGGAGTGGACCTTGGAGGCTGCCCCCAAGAT	212
QY	121	GGAGTCAGTGGGAACCTCCAGCGGCCGCCCTTCCGCCCCCACTCATCGAAAAACCT	180
DB	213	GGAGTCAGTGGGAACCTCCAGCGGCCGCCCTTCCGCCCCCACTCATCGAAAAACCT	272
QY	181	GCTCAGCGCGCGCGCGGAGCCATGAGAGCCTTCCGGATGCTGCTACTCAAAAGCA	240
DB	273	GCTCAGCGCGCGCGCGGAGCCATGAGAGCCTTCCGGATGCTGCTACTCAAAAGCA	332
QY	241	CCTCGCTGACATTCCACTGGAAGCTTTGGGGGCGCCACCGGGCGCGGGGGGCTCG	300
DB	333	CCTCGCTGACATTCCACTGGAAGCTTTGGGGGCGCCACCGGGCGCGGGGGGCTCG	392
QY	301	CACACCCGAAGAACCATCTTTCCACCCAGCAAGGGGGTGCAGCGCCACAGGTGCCATCCC	360
DB	393	CACACCCGAAGAACCATCTTTCCACCCAGCAAGGGGGTGCAGCGCCACAGGTGCCATCCC	452
QY	361	CCTGTTGCTGTTGACTCCCCCGGGGGCCACCTCCACCGCGCTGGGTCTGCTAGGTG	420
DB	453	CCTGTTGCTGTTGACTCCCCCGGGGGCCACCTCCACCGCGCTGGGTCTGCTAGGTG	512
QY	421	CTCTCATGCTGAGGATGGGTGAGAGGTCTCCACAGTGCCTCTGGGCGCCCCCATGG	480
DB	513	CTCTCATGCTGAGGATGGGTGAGAGGTCTCCACAGTGCCTCTGGGCGCCCCCATGG	572
QY	481	AGGAAGTGGACTCAGGTGGACTCCAAAGTCTCCTTGACCCCTGACTCGGGCCTCCTTT	540
DB	573	AGGAAGTGGACTCAGGTGGACTCCAAAGTCTCCTTGACCCCTGACTCGGGCCTCCTTT	632
QY	541	CATGTACTCTGCCACAGCTTTGGGGGACAACTCTGGGCGAGAGGGGAGCGCA-CTTGG	599
DB	633	CATGTACTCTGCCACAGCTTTGGGGGACAACTCTGGGCGAGAGGGGAGCGCACTTGG	692
QY	600	CACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGACAGTCTGGGGGACCATGTGGCC	659
DB	693	CACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGACAGTCTGGGGGACCATGTGGCC	752
QY	660	AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAAGAGGCGAGAGGCCCTGGGGAGA	719
DB	753	AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAAGAGGCGAGAGGCCCTGGGGAGA	812
QY	720	AGCCGGCCAGACAGACCCCTTCCGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGG	779
DB	813	AGCCGGCCAGACAGACCCCTTCCGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGG	872
QY	780	ACGAATCTCTCAAGCTATGCGAGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA	839
DB	873	ACGAATCTCTCAAGCTATGCGAGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA	932
QY	840	AGCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGCCCTGGTGTGC	899
DB	933	AGCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGCCCTGGTGTGC	992
QY	900	AGCTGATCCAGTCTTACCGAGGATGCCAGGCAATGCCATGTGAGGGGCTTCCGAGTGG	959
DB	993	AGCTGATCCAGTCTTACCGAGGATGCCAGGCAATGCCATGTGAGGGGCTTCCGAGTGG	1052

QY	960	CTTATAAGCGGCACGCTGCTGACCATGGATGACTTGGGGACCTTGTATGGACAGAATGGC	1019
DB	1053	CTTATAAGCGGCACGCTGCTGACCATGGATGACTTGGGGACCTTGTATGGACAGAATGGC	1112
QY	1020	TCAATGACAGGTGATGAACATGTATGGAGACCTGGTCAATGACAGAGTCCCTGAAAGG	1079
DB	1113	TCAATGACAGGTGATGAACATGTATGGAGACCTGGTCAATGACAGAGTCCCTGAAAGG	1172
QY	1080	TGCATTTCTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGTTTATGATGGGTGA	1139
DB	1173	TGCATTTCTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGTTTATGATGGGTGA	1232
QY	1140	AAAGTGGACCAAAACACGTGGACATCTTCAATAAGGAGCTACTGCTAAATCCCATCCACC	1199
DB	1233	AAAGTGGACCAAAACACGTGGACATCTTCAATAAGGAGCTACTGCTAAATCCCATCCACC	1292
QY	1200	TGGAGTGCATTTGGTCCCTCATCTCTGTTGATGTGTAGGGAGCGACCATCACTATTTTG	1259
DB	1293	TGGAGTGCATTTGGTCCCTCATCTCTGTTGATGTGTAGGGAGCGACCATCACTATTTTG	1352
QY	1260	ACTCGCAGCTACCCCTAAACCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAG	1319
DB	1353	ACTCGCAGCTACCCCTAAACCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAG	1412
QY	1320	AGCGGTAAGAAAGAACCGACTGGATTCCACAGGGCTGGAAAGGTTTACTTCAAAATGA	1379
DB	1413	AGCGGTAAGAAAGAACCGACTGGATTCCACAGGGCTGGAAAGGTTTACTTCAAAAT--	1470
QY	1380	ATGTGCCAGGCAGAAATATGACAGTGTGGTGTCTTTGTTGTGAGTACTGCAAGC	1439
DB	1471	-----GACTGCAAGC	1481
QY	1440	ATCTGCCCTGCTCAGCCATTACAGTTCACCCAGCAGACATGCCCAACATCTCGCGC	1499
DB	1482	ATCTGCCCTGCTCAGCCATTACAGTTCACCCAGCAGACATGCCCAACATCTCGCGC	1541
QY	1500	AGATCTACAAGAGCTGTGCTCACTGCAAACTCAGTGTGAGCCTCGTACCCAGACCC	1559
DB	1542	AGATCTACAAGAGCTGTGCTCACTGCAAACTCAGTGTGAGCCTCGTACCCAGACCC	1601
QY	1560	AGCCCATAAATGGGAGGAGACATGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC	1619
DB	1602	AGCCCATAAATGGGAGGAGACATGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC	1661
QY	1620	CTCTCTGCTCTTCCCTCAGTTCCTCTGTTGAGAGGAGTGGCCCTGTGGCTGGTGAGCA	1679
DB	1662	CTCTCTGCTCTTCCCTCAGTTCCTCTGTTGAGAGGAGTGGCCCTGTGGCTGGTGAGCA	1721
QY	1680	TGTATTTTCTTTTGTGAGAGAACTTGTGTTGATTTCTGATGTGAGGGGTGGCTACA	1739
DB	1722	TGTATTTTCTTTTGTGAGAGAACTTGTGTTGATTTCTGATGTGAGGGGTGGCTACA	1781
QY	1740	GAAAGCCCTTCTTCTCTCTGTTGAGAGGAGTGGCCCTGTGGCTGGGTGGAGCA	1799
DB	1782	GAAAGCCCTTCTTCTCTCTGTTGAGAGGAGTGGCCCTGTGGCTGGGTGGAGCA	1841
QY	1800	GTCTCTCTCCCTTCCCTGTCAGGAGCAGGAATCAGTGTGGGGGTGGTGGCGGA	1859
DB	1842	GTCTCTCTCCCTTCCCTGTCAGGAGCAGGAATCAGTGTGGGGGTGGTGGCGGA	1901
QY	1860	CAATAGGATCACTGCTGCTCAGATCTTCAAACTTTTATATATATATATATATATAT	1919
DB	1902	CAATAGGATCACTGCTGCTCAGATCTTCAAACTTTTATATATATATATATATATAT	1941
QY	1920	AT	1979
DB	1942	AT	2001
QY	1980	TCATAAAG	1988
DB	2002	TCATAAAG	2010







CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed specification.

XX

SQ Sequence 1804 BP: 431 A: 477 C: 468 G: 428 T: 0 other:

### Query Match

Query Match 78.8%; Score 1568.2; DB 22; Length 1804;

Best Local Similarity 97.18; pred. No. 0; pred. No. 0;

Best local similarity	97.1%	Area NO: 0
Matches 1719;	Conservative	0; Mismatches
28;	Indels	24; Gaps
11:		

[illegible]

Ov 234 AAAAGCACCTCGCTGACATTCCACTGGGAAGCTTTGGGGGGCCACCGGGGGCGGGCGG 293

[illegible]

Db 1793 AAAGCACCTCGCTGACATTCACCTGGAGCTTGGGGGGCCACCGGGCGGGCGGG 1734

52 1755 AATTGTCACCTCCCTGACATATCCACTGGAATGCTTTGGGGGCGCCACCGGGCCCGCGCGCG I/34

294 GGCCTCGCACACCCCAAGAACCATCTTTTACCCACGACAGGGGTGGACGCGACACAGCTC 353

QY Z54 98CCICGACACCCCHAGHACCACTTACCCCHAGCAGGGGTGCGACGCCACAGTG 333

db 1733 GGCTTCGCACACCCCAAGAACCACTCTTCAACCCAGCAAGCGCTCCCGCCACACCTC 1674

DD 1733 GGCCCTCGCACACCCCAAGAACCATCTTTACCCAGCAAGGGGGTGGGACGCCACAGGTG 1674

[illegible]

QY 354 CCATCCCCCTGTGTGTCGTATTGACTCCCCCGGGGGCCACCTCCACCCCCTGGTCTG 413

[illegible]

Db 1673 CCATCCCCCTGTTGTCGTTTGGACTCCCCCGGGGGCCACCTCCACCCCGGCTGGGCTG 1614

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

QY 414 CTAGGTGCTCTCATGGCTGAGGATGGGGTGAGAGGGTCTCCACCAGTGCCCTCTGGGCC 473

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Db 1613 CTAGGTGCTCTCATGGCTGAGGATGGGGTGAGGGTCTCCACCATGCCCCCTCTGGGCC 1554

Qy 474 CCCATGGAGGAAGATGGACTCAGGTGGACTCCAAGTCTCCTCTGGACCCCTGACTCGGC 533

\_\_\_\_\_

Db 1553 CCCATGGAGGAAGATGGACTCAGGTGGACTCCAAGTCTCCTCTGGACCCTGACTCGGC 1494

---

QY 534 CTCCTTTCATGTACTCTGCCCAACGGTTTGGGGGACAATCTGGGCCAGAAAGGGAGCGC 593

\_\_\_\_\_

Db 1493 CTCCTTTCATGTA CTCTGCCCAACGGTTT TGGGGGACAATCTGGGCCAGAACGGAGCGC 1434

[illegible]

OV 594 A-CTTGGCACCCCCCTGATGCCAGCATCCTCATCAGCAATGTGTGCAGCATCGGGACCAT 652

[illegible]

Db 1433 AGCTTGGCACCCTGATGCCAGCATCCTCATCAGCAATGTGTGCAGCATCGGGGACCAT 1374

DD 1433 AGCTTGGCACCCCTGATGCGCAGCATCCCTCATCAGCAATGTGTGCGAGCATCGGGGACCAT 1374

653 CTTGGCCCAAGGACCGCTTCTTCACGCCGCATCAGCATTCCTCACACCCACTC

Ov 712

inflammation; neurological disorder; Alzheimer's disease; food additive; angioygenic disorder; kidney disorder; gastrointestinal disorder; allergy; pregnancy-related disorder; endocrine disorder; infection; wound healing; cell culture; chemotaxis; vulnery; binding partner identification; gene therapy; ss.

Homo sapiens.

Key Location/Qualifiers  
CDS 441..713  
/tag= a  
/product= "Human secreted protein precursor"  
sig\_peptide 441..521  
/tag= b  
mat\_peptide 522..710  
/tag= c  
/product= "Mature human secreted protein"

WO200134767-A2.

17-MAY-2001.

01-NOV-2000; 2000WO-US30036.

05-NOV-1999; 99US-0163576.

27-JUL-2000; 2000US-0221366.

(HUMA-) HUMAN GENOME SCI INC.

Soppet DR, Komatsoulis G, Shi Y, Olsen HS, Ruben SM;

WPI; 2001-316492/33.

P-PSDB; AAE01682.

Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition - Claim 1; Page 417-418; 540pp: English.

AAD05492-AAD05564 represent cDNAs corresponding to 22 human secreted protein genes, and AAE01672-AAE01743 represent the proteins they encode. AAE01744-AAE01763 represent human secreted protein fragments or variants. The secreted proteins and their genes are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 22 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g. rheumatoid arthritis), inflammation, allergies, autoimmune neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angioygenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.

Sequence 1934 BP; 423 A; 542 C; 551 G; 418 T; 0 other;

Query Match 72.9%; Score 1451.2; DB 22; Length 1934;  
Best Local Similarity 89.2%; Pred. No. 0;

		Matches 1672; Conservative 0; Mismatches 178; Indels 24; Gaps 9;			
QY	57	GAGGAGGAGGAGGAGATGAAGATGAAGAGGAGGAGAGTGCAGCTGGAGGCTGCCCA	116		
DB	12	GAAGAGGAGGATGAGGATGAAGATGAAGAGGAGGAGTGCAGCTGGAGGCTGCCCA	71		
QY	117	AGATGAGTCACTGGGAAACCTCCACGCGCCCGCCCTTCCGCGCCCACTCATCGAAA	176		
DB	72	AGATGGGTGAGTGGAGCTTCCACAGCGCCCTCGCTCCTCCGCGCCCTACTCATCGAAA	131		
QY	177	ACCTGCTACAGCGCGCGCGAGCCATGAGAGCCTTCGGATGCTGCTACTCAAAA	236		
DB	132	ACCTGCTCAGCGCGCGCGAGCCATGAGAGCCTTCGGATGCTGCTACTCAAAA	191		
QY	237	AGCACCTCGCTGACATTCACCTGGAGCTTGGGGCGCCACCGGGCGCGGGCGGGG	296		
DB	192	AGCACCTCGCTGACATTCACCTGGAGCTTGGGGCGCCACCGGGCGCGGGCGGGG	250		
QY	297	CTCGCACACCCCAAGAACCATCTTTTCCACCCACCAAGAGGGGTCCGACGACACAGSTGCCA	356		
DB	251	CTCGCACACCCCAAGAACCATCTTTTCCACCCACCAAGAGGGGTCCGACGACACCGAGSTGCCA	310		
QY	357	TCCCTCTGTTGCTGTTTGAATTCCTCCCGCGGGGCGCCACCTCCACCGCGGTGGGTCTGCTA	416		
DB	311	TCCCTCTGTTGCTGTTTGAATTCCTCCCGCGGGGCGCCACCTCCACCGCGGTGGGTCTGCTA	370		
QY	417	GGTGTCTCATGCTGAGGATGGGTGAGAGGCTCTCCACAGTGCCTCTCTGGGGCCCC	476		
DB	371	GGTGTCTCATGCTGAGGATGGGTGAGAGGCTCTCCACAGTGCCTCTCTGGGGCCCC	430		
QY	477	ATGGAGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTTGACCTGACTCGGGCTC	536		
DB	431	ATGGAGGAAGATGGATTACGTTGGACTCCAAAGTCTCTCTTGACCTGACTCGGGCTC	490		
QY	537	CTTTTCAATGACTGCTGCCCAACGGTTTTGGGGGACAAATCTGGGCGAGAGGGAGCCACT	596		
DB	491	CTCTCTGTACTCTTCCCAATGGCTTTGGGGGACCCCGCGGGCGAGAGGGGGGT	550		
QY	597	-TGGCACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGAGCATCGGGGACCATGTG	655		
DB	551	CTGGCACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGAGCATCGGGGACCATGTG	610		
QY	656	GCCAGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGGAGAGGCGCTGGG	715		
DB	611	GCCAGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGAGGAGAGGCGCTGGG	670		
QY	716	GAGAAAGCGCGCGAGCAGAGCCCTGCGAGAGGAGCATGTGACCTGCCTACAGAGCATC	775		
DB	671	GAGAAAGCGCGCGAGCAGAGCCCTGCGGGAGGAGCATGTGACCTGCCTACAGAGCATC	730		
QY	776	TTGGAGGAATTCCTTCAACGATATGCGACGCTCATACCCCTCAGCACTCATGAGGTAGTA	835		
DB	731	TTGGATGAATTCCTTCAACGATATGCGACGCTCATACCCCTCAGCACTCATGAGGTAGTA	790		
QY	836	GAGAGCTGGAGGACATTTTCCAGGAGGAGTTTTCACCCCTTCCAGGAGGCGCTGGTG	895		
DB	791	GAGAAATTAGAGACATTTTCCAGGAGGAGTTCTCTACCTTCCAGGAGGCGCTGGTG	850		
QY	896	TTGAGCTGATCAGTCTTACAGGAGGATGCCAGGCAATGCCATGGTGGAGGCTTCCGA	955		
DB	851	CTGAGCTGATCAGTCTTACAGGAGGATGCCAGGCAATGCCATGGTGGAGGCTTCCGA	910		
QY	956	GTGGCTTAAAGCGGACGCTGACCATGGATGACTTGGGACCTTGTATGAGACAGAAC	1015		
DB	911	GTGGCTTAAAGCGGACGCTGACCATGGATGACTTGGGACCTTGTATGAGACAGAAC	970		
QY	1016	TGGCTCAATGACGAGTGAATGAACATGTATGGAGAGCTGGTGCATGGACAGTCCCTGAA	1075		
DB	971	TGGCTCAACGACGAGTGAATGAACATGTATGGAGAGCTGGTGCATGGACAGTCCCTGAA	1030		
QY	1076	AAGGTGCATTTCTTCAATAGTTTCTTCTATGATAAAGTCCGTTACCAAGGGTTATGATGG	1135		
DB	1031	AAGGTGCATTTCTTCAACAGTTTCTTCTACGATAAGTCCGTTACCAAGGGTTACGATGGA	1090		





```
Db 72 AGATGGGGTCAGCTGGGAGCTTCCACGGGCTCGTCTCCCGCCCTACTCATCGAAAA 131
Qy 177 ACCTGCTCAGACGCCCGCCGAGCCATGAGAGCCTTCGGAGTCTGCTACTCAAAA 236
Db 132 ACCTGCTCCAGCGCCGCGGAGCCATGAGAGCCTTCGGATGCTGCTACTCGAAA 191
Qy 237 AGCACCTCGCTGACATTCCTCACTGGAAGCTTTGGGGGCCCCACCGGGCCCGCGCGGGC 296
Db 192 AGCACCTCGCTGACATTCCTCACTGGAAGCTTTGGGGGCCCCACCGGGCCCGCGCGGAG- 250
Qy 297 CTGCGACACCCCAAGAACCATCTTTCACCCCAAGCAAGGGGGTGGAGCCACAGGTGCCA 356
Db 251 CTGCGACACCCCAAGAACCATCTTTCACCCCAAGCAAGGGGGTGGAGCCACAGGTGCCA 310
Qy 357 TCCCCCTGTTGCTGCTTTTGACTCCCGCCGGGGCCACTCTCACCCGGCTGGGTCTGCTA 416
Db 311 TCCCCCTGCTGCTTTTGACTCCCGCCGGGGCCACTCTCACCCGGCTGGGTCTGCTA 370
Qy 417 GGTGCTCTCATGGCTGAGGATGGGGTGAGAGGCTCTCCACAGTGCCTCTGGACCTGACTCGGGCCTC 536
Db 371 GGTGCTCTCATGGCTGAGGATGGGATGAGAGGCTCTCCACCGATGCTCTGGGCCCTC 490
Qy 537 CTTTTCATGTACTCTGCCCCAACGGTTTTTGGGGGCAATCTGGGCCAGAAAGGGAGCGCACT 596
Db 491 CTCTCTGTACTCTTCCCATGAGCTTTGGGGGACCCCGCCGGCCAGAAAGGGAGCGGGGT 550
Qy 597 -TGGCACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGCAGCATCGGGGACCATGTG 655
Db 551 CTGSCACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGCAGCATCGGGGACCATGTG 610
Qy 656 GCCCAGAGCTTTTCAGGCTCAGATTTGGGCATGTCAGCAAGAGGAGAGAGGCTGGG 715
Db 611 GCCCAGAGCTATTTTCAGGCTCAGATCTGGGCACCCAGAGAGGAGGAGCGGCC-GGG 669
Qy 716 GAGAAAGCCGGCAGCAGACCCCTCGAGAGAGGAGCATGTGACCTCGCTACAGAGCATC 775
Db 670 GAGAAAGCCGGCAGCAGACCCCTCGGGAGGAGCATGTGACCTCGCTGAGAGCATC 729
Qy 776 TTGACGAATTCCTTCAAGCATATGCGAGCCTCATACCCCTCAGCATGATGAGGTAGTA 835
Db 730 TTGATGAATTCCTTCAAGCATATGCGAGCCTCATACCCCTCAGCAGGACGAGGTAGTA 789
Qy 836 GAGAGCTGAGAGCATTTTCCAGCAGGAGTTTTCACACCCCTTCCAGGAAGGCGCTGGTG 895
Db 790 GAGAAATTAGAGGACATTTTCCAGCAGGAGTTTCTTACACCTTCCAGGAAGGCGCTGGTG 849
Qy 896 TTGACGCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGTTGAGGGGCTTCCGA 955
Db 850 CTGACGCTGATCAGTGTACAGGGATGCCAGGCAACCCATGTTGAGGGGCTTCCGG 909
Qy 956 GTGGCTTATAGCGGCACGTGTCAGCATGATGACTTTGGGGACCTTGTATGGACAGAAC 1015
Db 910 GTGGCTTACAGCGGCACGTGTCAGCATGATGACTTTGGGCACCTTATATGGACAGAAC 969
Qy 1016 TGGCTCAATGACAGGTGATGACATGTATGGAGACCTGTGATGACACAGTCCCTGAA 1075
Db 970 TGGCTCAACAGGTGATGACATGTACGGAGACCTGTGATGACAGCGTCCCGGAG 1029
Qy 1076 AAGGTGATTTCTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGTTTATGATGG 1135
Db 1030 AAGGTGATTTCTCAACAGTTTCTTCTACGATAAGCTCGGTACCAAGGTTTACGATGGA 1089
Qy 1136 GTGAAAAGGTGGACCAAAAACGTGGACATCTTCAATAAGGAGTACTGCTTAATCCCATC 1195
Db 1090 GTGAAAAGGTGGACCAAAAACGTGGACATCTTCAATAAGGAGTCTCTGCTTAATCCCATC 1149
Qy 1196 CACCTGGAGTGTATGGTCCCTCATCTCTGTTGATGTGAGGCGGACGCACCATCATCTAT 1255
Db 1150 CACCTGGAGTGTACCTGGTCCCTCATCTCTGTTGATGTGAGGCGGCGCACCATCATCTAT 1209
```

```
Qy 1256 TTTGACTCGCAGCGTACCTTAAACCCGCTGCCCTAAGCATATTGCCAAGTATCTACAG 1315
Db 1210 TTGACTCGCAGCGCACCTTAAACCCGCTGCCCTAAGCATATTGCCAAGTATCTACAG 1269
Qy 1316 GCAGAGCGGTAAAGAAAGACCGACTGGATTTCCACAGGCGTGGAAAGGTTACTTCAA 1375
Db 1270 GCAGAGCGACTGAAGAAAGACCGGCTGGATTTCCACAGGCGTGGAAAGGTTACTTCAA 1329
Qy 1376 ATGAATGTGCCAGGCGAGATATGACAGTACTGTGGTCTTTTGTGTTGCGAGTACTGC 1435
Db 1330 ATGAATGTGCCAGGCGAGATATGACAGTACTGTGGTGGCTTTGTGTTACAGTACTGC 1389
Qy 1436 AAGCATCTGCCCTCTGCTCAGCCATTACGCTTCCACCCAGCAGGACATGCCCAAACTTCGT 1495
Db 1390 AAGCATCTGCCCTCTGCTCAGCCATTACGCTTCCACCCAGCAGGACATGCCCAAACTTCGC 1449
Qy 1496 CGCAGATCTACAGAGCTGTCTCACTGCAAACTCACTGTGTGAGCTGTACCCAGAG 1555
Db 1450 CGCAGATCTACAGAGCTGTCTCACTGCAAACTCACTGTGTGAGCTGTATCCAGG 1509
Qy 1556 CCCCAGCCCA----TAAATGGGAAGGAGACATGGAG-----TCCCTTCCCAGAACT 1607
Db 1510 CTTCAAGCCCATTTGGTCAATGGGACATGGGGGACCTTCTTCCCAGAACT 1569
Qy 1608 CCAAGTTCCCTTTCCTCTTCCCTCTTCCCACTCACTTCCCTTTGGTTTTCATATTTAA 1667
Db 1570 CCAAGTTCCCTCTTTCCTCTCC-----CCCTCCAGTTCCTATGTTTTCATATTTAA 1625
Qy 1668 TGTTTCA----ATTTCTGTAATTTTTTTTCTTTGAGAGAACTTGTGATTTCTGATGT 1723
Db 1626 TGTTTAAATGAATTTCTGATATTTTTTTTCTTTGAGAGAACTTGTGATTTTGTATGT 1685
Qy 1724 GCAGGGGTGGCTACAGAAAAGCCCTTCTTCTCTGTTTGCAGGGGAGTGTGGCCCTG 1783
Db 1686 TCAGGGGTGGCTACAGAAAGCCCTT--CTCCCTCTGCTAGGAGGAGTGTGGCCCTG 1743
Qy 1784 TGCCCTGGGTGGAGCAGTCAAT---CTCCCTTCCCTTCCCTGTCAGGAGCAGGAACTCAGT 1840
Db 1744 TGCCCTGGGTGGGCGGCTATCCCTTCCCTGTCAGGGGCGGGGAGGACATCTG 1803
Qy 1841 GCTGGGGTGGTGGGCGGACAATAGGATCACTGCCCTGCCAGATCTTCAAACTTTTATATA 1900
Db 1804 TGCTGGGTGGTGGGCGGGCAA-AGGGAGTGTGCTGCCAGATATTTCAAACTTTGTTTA 1862
Qy 1901 TATATATATATATA 1914
Db 1863 TAGTATATATAAA 1876

RESULT 9
AAD05541/c
ID AAD05541 standard; cDNA; 1507 BP.
XX
AC AAD05541;
XX
XX
DT 18-JUL-2001 (first entry)
XX
DE Human secreted protein-encoding gene 11 cDNA clone HOF0B27, SEQ ID NO:60.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angioneurotic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnery;
KW cell culture; chemotaxis; food additive;
KW binding partner identification; ss.
XX
OS Homo sapiens.
```







cell culture; chemotaxis; food additive;  
binding partner identification; ss.

**Homo sapiens.**

Key	Location/Qualifiers
CDS	27..167

/product= "Human secreted protein precursor"

sig\_peptide 27..158

SECRET

mat\_peptide 159..164 /\*tag= c

```

/  casg = 0
/ product =

```

11

WO200134767-A2.

17-MAY-2001

1007-1441

01-NOV-2000; 2000WO-US30036.

05-NOV-1999; 99US-0163576.  
27 THE 2000: 2000US 033135C

27-JUL-2007; 200005-0221366.

(HUMA-) HUMAN GENOME SCI INC.

.....

Soppet DR, Komat

WPI: 2001-316492/33.  
P-PSDB; AAE01720.

Claim 1; Page 446-447; 540pp; English.

AA05432-AD05564 represent cDNAs corresponding to 22 human secreted protein genes, and AA051672-AA051743 represent the proteins they encode. AA051744-AA051763 represent human secreted protein fragments or variants. The secreted proteins and their genes are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 22 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic diseases, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.

Sequence 695 BP: 131 A: 224 C: 217 G: 117 T: 6 other:

ery Match	27 78.	Score 551.	DB 22.	Length 695.
-----------	--------	------------	--------	-------------

st Local Similarity 90.28; Pred. No. 2.2e-119;

tches 617; Conservative 5; Mismatches 56; Indels 6; Gaps 3;

57 GAGGAGGAGGAGGATGAAGATGAAGAGGAGGAGTGGAGGCTGCCCCCA 116

Db	17	GAAGAGAGGATGAGGATGAAGATGAGAAAGAGAAAGTGGCAGCTTGGAGGCTGGCCCCC	76
Qy	117	AGATGGAGTCAGCTGGGAACCTCCACAGCGCCCCGCTTCCCGCCCCCACTCATCGAAAA	176
Db	77	AGATGGGTCAGCTGGGAGCTTCCACAGGCGCTCTGCTCTCCCGCCCTACTCATCGAAA	136
Qy	177	ACCTGCTCACAGCGCCGCCCGAGCCATGAGAGCCTTCGGATGTGCTCTACTCAAA	236
Db	137	ACCTGCTCCAGCGCCGCCGAGCCATGAGAGCCTTCCGATGCTGTACTCTCGAA	196
Qy	237	AGCACCCTGCCTGACATTCACACTGGAAGCTTTGGGGGGCCACCGGGCGCGCGCGGCG	296
Db	197	AGCACCCTGCCTTAACATTCACACTGGAAGCTTTGGCGCGCCACCGGGCGCGCGCGGAG-	255
Qy	297	CTCGCACACCCCAAGAACCATCTTTCACCCACGAAAGGGGTGCGACGCCACAGGTGCCA	356
Db	256	CTCGCACACCCCAAGAACCATCTTTCACCCACGAAAGGGGTGCGACCGCAGGTGCCA	315
Qy	357	TCCCCCTGTGTGCTTTTGACTTCCCGCCGGGGCCACTCCACCCCGGTGGGTCTGTCTA	416
Db	316	TCCCCCTGTGTGCTTTTGACTTCCCGCACGGGGCCACTCCACCCCGGTGGGTCTGTCTA	375
Qy	417	GGTCTCTCATGGCTGAGGATGGGTGAGAGGCTCTCCACCATGTGCCCTCTGGGCCCCC	476
Db	376	GGTCTCTCATGGCTGAGGATGGGTGAGAGGCTCTCCACCATGTGCCCTCTGGGCCCCC	435
Qy	477	ATGGAGGAAGATGACATCAGTGGACTCCAAAGTCTCTCTGACACCTGACTCGGGCCTC	536
Db	436	GTGGAGGAAGATGATATACGGTGGACTCCAAAGTCTCTCTGACACCTGACTCGGGCCTC	495
Qy	537	CTTTTCATGTACTTCGCCCAACGGTTTGGGGGCAATCTGGGGCCAGAAAGGAGCGCACT	596
Db	496	CTCTCCTGTACTCTTCCCAATGGCTTTGGGGGACCCCGCGGCCAGAAAGGAGCGGGT	555
Qy	597	-TGGACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGACAGCATCGGGGACCATGTG	655
Db	556	CTGGCAGCCCCGTGATGCCAGCATCCTCATCAGCAATGTGTGACAGCATCGGGGACCATGTG	615
Qy	656	GCCCAGAGGCTTTTTCAGGGCTCAGATTGGGATGGCAGCAAGAGGAGCGGCTGGG	715
Db	616	GCCCAGAGGCTTTTTCAGGGCTCAGATTGGGATGGCAGCAAGAGGAGCGGCTGGG	671
Qy	716	GAGAAAGCGGCCACAGCAGCCCC	739
Db	672	GAGAAAGCGGCCACAGCAGCCCC	695
RESULT	12		
AAAF5961	ID	AAAF5961 standard; DNA: 5318 BP.	
XX	AC	AAAF5961;	
XX	DT	12-APR-2001 (first entry)	
XX	DE	Human eukaryotic initiation factor 4A1 promoter fragment.	
XX	KW	Human; eukaryotic transcription factor 4A; eIF4A1; promoter;	
XX	KW	DNA construct; DNA vaccine; gene therapy; cancer; Parkinson's disease;	
XX	OS	cystic fibrosis; hepatitis; HIV; ds.	
XX	FN	Homo sapiens.	
XX	XX	WO200102594-A2.	
XX	XX	11-JAN-2001.	
XX	PF	05-JUL-2000; 2000WO-GB02569.	
XX	PR	06-JUL-1999; 99GB-0015638.	
XX	PR	14-DEC-1999; 99GB-0029547.	

```
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Greaves DR, Thomsen L, Catchpole IR, Ford MJ;
XX
XX WPI; 2001-138151/14.
XX
PT New DNA constructs, useful in gene therapy for treating or preventing
PT chronic inflammatory reactions such as multiple sclerosis or asthma, as
PT DNA vaccines, or in the commercial production of proteins -
XX
XX
PS Claim 19; Page 58-62; 62pp; English.
XX
CC The present invention describes a DNA construct comprising a
CC transpositional regulatory sequence linked to the gene of interest where
CC the regulatory sequence is the human eukaryotic initiation factor 4A1
CC (eIF4A1) gene promoter. This can be used in the production of DNA
CC vaccines and gene therapy, both of which can be used in the prevention
CC and treatment of diseases such as cystic fibrosis, Parkinson's disease,
CC cancer, hepatitis B and C, HIV, tuberculosis, HPV, HSV, multiple
CC sclerosis, Alzheimer's disease and asthma. It can also be used in
CC vaccines for contraception or drug addition.
XX
SQ Sequence 5318 BP; 1327 A; 1210 C; 1423 G; 1350 T; 8 other;

Query Match      24.4%; Score 486.4; DB 22; Length 5318;
Best Local Similarity 94.9%; Pred. No. 7.6e-104;
Matches 544; Conservative 1; Mismatches 17; Indels 11; Gaps 4;

QY 1427 CAGTACTGCAAGCATCTGCCCTGTCTCAGCCATTTCAGCTTCACCCAGGACATGCCCC 1486
DB 1427 CAGTACTGCAAGCATCTGCCCTGTCTCAGCCATTTCAGCTTCACCCAGGACATGCCCC 3886
QY 1487 AAATCTTCGTCGGCAGATCTACAAAGAGCTGTGTCACTGCAAACTCACTGTGTGAGCTCG 1546
DB 1487 AAATCTTCGTCGGCAGATCTACAAAGAGCTGTGTCACTGCAAACTCACTGTGTGAGCTCG 3946
QY 1547 TACCCAGAGCCCCAAGCCATAAATGGGAGGAGACATGGGAGTCCCTTCCCAAGAAC 1606
DB 1547 TACCCAGAGCCCCAAGCCATAAATGGGAGGAGACATGGGAGTCCCTTCCCAAGAAC 4006
QY 1607 TCCAGTTCCTTCTCTCTGTGCTCTTCCACATCACTTCCCTTGTGTTTTCATATTTAA 1666
DB 1607 TCCAGTTCCTTCTCTCTGTGCTCTTCCACATCACTTCCCTTGTGTTTTCATATTTAA 4066
QY 1667 ATGTTTTCAATTTTCTGATTTTCTTTTCTTGAGAGAACTACTTGTGATTTCTGATGTGC 1725
DB 1667 ATGTTTTCAATTTTCTGATTTTCTTTTCTTGAGAGAACTACTTGTGATTTCTGATGTGC 4126
QY 1726 AGGGGTGGCTACAGAAAAGCCCTTCTCTCTCTTTTGCAGGGAGTGTGCCCTGTG 1785
DB 1726 AGGGGTGGCTACAGAAAAGCCCTTCTCTCTCTTTTGCAGGGAGTGTGCCCTGTG 4186
QY 1786 G-CCTGGTGGAGCAGTCACTCTCCCTTCCCTTCCCTGGGAGGAGCAATCAGTGCN- 1843
DB 1786 G-CCTGGTGGAGCAGTCACTCTCCCTTCCCTTCCCTGGGAGGAGCAATCAGTGCN- 4246
QY 1844 GGGGTGGTGGGGGACAAATAGGATCACTGCTTCCAGATCTTCAAACCTTTTATATATAT 1903
DB 1844 GGGGTGGTGGGGGACAAATAGGATCACTGCTTCCAGATCTTCAAACCTTTTATATATAT 4306
QY 1904 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1957
DB 1904 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 4366
QY 1958 --AAATGCCAGGCTGCTGTGTCATTAAG 1988
DB 1958 --AAATGCCAGGCTGCTGTGTCATTAAG 4399

RESULT 13
AAI94288/c
ID AAI94288 standard; cDNA; 870 BP.
XX
```

```
AC AAI94288;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 363.
XX
KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO200166719-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-JP01629.
XX
PR 07-MAR-2000; 2000JP-0159195.
XX
PP (CHIB-) CHIBA PREFECTURE.
PA (HISM ) HISAMITSU PHARM CO LTD.
XX
PI Nakagawara A;
XX
XX WPI; 2001-565584/63.
DR Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents -
XX
PS Claim 1; Page 308-309; 2979pp; Japanese.
XX
CC The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.
XX
SQ Sequence 870 BP; 217 A; 229 C; 190 G; 211 T; 23 other;

Query Match      23.5%; Score 458.4; DB 22; Length 870;
Best Local Similarity 94.3%; Pred. No. 5.9e-100;
Matches 533; Conservative 0; Mismatches 27; Indels 5; Gaps 5;

QY 519 GACCCCTGACTCGGCGCTCCTTTCATGTACTCTGCCAACGGTTTGGGGACAATCTGGG 578
DB 519 GACCCCTGACTCGGCGCTCCTTTCATGTACTCTGCCAACGGTTTGGGGACAATCTGGG 810
QY 579 -CCAGAAGGGGAGCGCATTTGGCACCCCC-TGATGCCAGCATCCTCATCAGCAATGTGTG 636
DB 579 -CCAGAAGGGGAGCGCATTTGGCACCCCTTGCATGCCAGCATCCTCATCAGCAATGTGTG 750
QY 637 CAGCATCGGGGACCATGTGCCCCCAGGAGCTTTTTCAGGCTCAGATTGGGGCATGGCACA 696
DB 637 CAGCATCGGGGACCATGTGCCCCCAGGAGCTTTTTCAGGCTCAGATTGGGGCATGGCACA 692
QY 697 AGAGGCAGAGAGCGCTTGGGGAGAAAGCCGCCAGCACACGCCCTCGCAGAGAGGAGCATGT 756
DB 697 AGAGGCAGAGAGCGCTTGGGGAGAAAGCCGCCAGCACACGCCCTCGCAGAGAGGAGCATGT 752
QY 757 GACCTCGGTACAGAGCATC-TTGGAGCAATTCCTTAAACGATATGGCAGCCTCATACCCC 815
DB 757 GACCTCGGTACAGAGCATC-TTGGAGCAATTCCTTAAACGATATGGCAGCCTCATACCCC 572
QY 816 TCAGCACTGATGAGGTAGTAGAGAGCTGGAGACATTTTCCAGCAGGAGTGTTCACCCC 875
DB 816 TCAGCACTGATGAGGTAGTAGAGAGCTGGAGACATTTTCCAGCAGGAGTGTTCACCCC 512
QY 876 CTTCCAGGAAGGCGCTGTTGTCAGCTCATCCAGTCTTTACCAGCGGATGCCAGGCAATG 935
DB 876 CTTCCAGGAAGGCGCTGTTGTCAGCTCATCCAGTCTTTACCAGCGGATGCCAGGCAATG 452
```

QY 936 CCATGGTGAAGGGCTTCCGAGTGGCTTATAAGCGGCACGTGCTGACCATGGATGACTTGG 995  
Db 451 CCATGGTGAAGGGCTTCCGAGTGGCTTATAAGCGGCACGTGCTGACCATGGATGACTTGG 392  
QY 996 GGACCTTGATGACAGAACTGGCTCAATGACAGGTGATGACATGATGATGAGACCTGG 1055  
Db 391 GGACCTTGATGACAGAACTGGCTCAATGACAGGTGATGACATGATGATGAGACCTGG 332  
QY 1056 TCATGGACACAGTCCCTGAAAAGGT 1080  
Db 331 TCATGGACACAGTCCCTGAAAAGGT 307  
RESULT 14  
ID AAS25529/c  
XX AAS25529 standard; cDNA; 503 BP.  
AC AAS25529;  
DT 07-NOV-2001 (first entry)  
XX Human ovarian PCR-subtracted cDNA library clone #1614.  
DE  
XX Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;  
KW gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;  
KW primer; probe.  
OS Homo sapiens.  
XX  
PN WO200157207-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 05-FEB-2001; 2001WO-US03733.  
XX  
PR 04-FEB-2000; 2000US-0180403.  
PR 28-MAR-2000; 2000US-0192745.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Algate PA, Mannion J;  
XX  
DR WPI; 2001-488879/53.  
XX  
PS New polynucleotides encoding ovarian tumour proteins, useful for  
PT treating ovarian cancer, and as probes, primers, and markers of cancer  
PT progression -  
XX  
PS Example 1; page 375; 378pp; English.  
XX  
CC The invention comprises compositions used for the therapy and diagnosis  
CC of ovarian cancer. The compositions comprise one or more ovarian tumour  
CC proteins, their associated polynucleotides, or immunogenic portions of  
CC the proteins. The ovarian tumour polynucleotides and polypeptides are  
CC useful for stimulating and/or expanding T cells specific for a tumour  
CC protein. They are also useful for inhibiting the development of cancer in  
CC a patient with an ovarian tumour DNA or protein by incubating isolated  
CC T-cells allowing them to proliferate, and administering to the patient.  
CC The sequences can be used as markers for cancer, for example, to monitor  
CC ovarian cancer progression. Probes and primers are useful in nucleic acid  
CC hybridisation, in detecting the presence of complementary sequences in a  
CC given sample, for preparing mutant species and for preparing other  
CC genetic constructions. Sequences AAS23820-AAS25231 and AAS25528-AAS25549  
CC represent human ovarian tumour protein cDNA clones.  
SQ Sequence 503 BP; 120 A; 110 C; 147 G; 126 T; 0 other;  
XX  
Query Match 22.2%; Score 442; DB 22; Length 503;  
Best Local Similarity 90.8%; Pred No. 7.3e-94;  
Matches 503; Conservative 0; Mismatches 0; Indels 51; Gaps 1;  
QY 1111 ACTCCGTACCAAGGGTTATGATGGGTGAAAAGGTGGACCAAAACGTTGGACATCTTCAA 1170  
|||||

Db 503 ACTCCGTACCAAGGGTTATGATGGGTGAAAAGGTGGACCAAAACGTTGGACATCTTCAA 444  
QY 1171 TAAGGAGCTACTGCTAATCCCATCCACCTGGAGGTGCATTGGTCCCTCATCTCTGTGA 1230  
|||||  
Db 443 TAAGGAGCTACTGCTAATCCCATCCACCTGGAGGTGCATTGGTCCCTCATCTCTGTGA 384  
|||||  
QY 1231 TGTGAGGCGACGACCATCACCCTATTTTGTGCTGCGAGCGTACCTTAACCGCGGTGCC 1290  
|||||  
Db 383 TGTGAGGCGACGACCATCACCCTATTTTGTGCTGCGAGCGTACCTTAACCGCGGTGCC 324  
|||||  
QY 1291 TAAGCATATTTGCCAAGTATCTACAGGCAGAGCGGTAAAGAACCGGACTGGATTCCA 1350  
|||||  
Db 323 TAAGCATATTTGCCAAGTATCTACAGGCAGAGCGGTAAAGAACCGGACTGGATTCCA 264  
|||||  
QY 1351 CCAGGCTGCAAGGTTACTTCAAAATGAATGTGGCCAGGCAGAAATAATGACAGTCACTG 1410  
|||||  
Db 263 CCAGGCTGCAAGGTTACTTCAAAAT-----GTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCACTTCAC 237  
|||||  
QY 1411 TGGTGCTTTTGTGTTGTCAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCACTTCAC 1470  
|||||  
Db 236 -----GTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCACTTCAC 195  
QY 1471 CCAGCAGGACATGCCCAAACTTCGTCGGCAGATCTACAAGGAGCTGTGCTGCAAACT 1530  
|||||  
Db 194 CCAGCAGGACATGCCCAAACTTCGTCGGCAGATCTACAAGGAGCTGTGCTGCAAACT 135  
QY 1531 CACTGTGTGAGCCTCGTACCCAGAGCCCAAGCCCAATTAATGGAGGAGGAGCATGGAG 1590  
|||||  
Db 134 CACTGTGTGAGCCTCGTACCCAGAGCCCAAGCCCAATTAATGGAGGAGGAGCATGGAG 75  
QY 1591 TCCCTTCCCAAGAACTCCAGTTCCTTCTCTTGGCTCTTCCCACTCACTTCCCTTT 1650  
|||||  
Db 74 TCCCTTCCCAAGAACTCCAGTTCCTTCTCTTGGCTCTTCCCACTCACTTCCCTTT 15  
QY 1651 GGTTCCTCATATTT 1664  
|||||  
Db 14 GGTTCCTCATATTT 1  
RESULT 15  
AAI93958/c  
ID AAI93958 standard; cDNA; 820 BP.  
XX  
AC AAI93958;  
XX  
DT 13-NOV-2001 (first entry)  
XX  
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 33.  
XX  
KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200166719-A1.  
PD 13-SEP-2001.  
XX  
PF 02-MAR-2001; 2001WO-JP01629.  
XX  
PR 07-MAR-2000; 2000JP-0159195.  
XX  
PA (CHIB-) CHIBA PREFECTURE.  
PA (HISM) HISAMITSU PHARM CO LTD.  
XX  
PI Nakagawara A;  
XX  
DR WPI; 2001-565584/63.  
XX  
PT Nucleic acids originating in gene expressed in human neuroblastoma,  
PT useful as probe or primer in diagnosing prognosis of human  
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker  
PT for anti-cancer agents -  
XX

PS Claim 1; Page 59; 2979pp; Japanese.

XX The invention relates to novel genes (AAI93926-AAI97963), expressed in  
CC human neuroblastoma. The nucleic acids are applicable as a probe or  
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and  
CC susceptibility indicators or tumour markers for anti-cancer agents. The  
CC gene information for diagnosing prognosis is related to factors similar  
CC to that for N-myc and TrkA genes.

XX  
SQ Sequence 820 BP; 224 A; 223 C; 170 G; 191 T; 12 other;

Query Match 18.2%; Score 362.2; DB 22; Length 820;  
Best Local Similarity 85.1%; Pred. NO. 4.7e-75;  
Matches 481; Conservative 0; Mismatches 70; Indels 14; Gaps 7;

QY	517	TGGACCTGACTCGGCGCTCTTTCATCTCTGTA	CTCTGCCAACGGTTTGGGGGACAATCTG	576
Db	820	TTGACCTTGATGGGCTCCNTT	NATGTTGCCAACGGTTTGGGGG---AAATTT	764
QY	577	GGCCAGAAGGGGAGCGCAGCTTGGCACCCCTG	ATGCCAGCATCCTCATCAGCAATGTGTG	636
Db	763	GGCCANAGGGGACGAGTTTGACCCCTTGATCCA	--GATCCTTATCAGCAATNTTT- 707	
QY	637	CAGCATCGGGGACCATGTGCCCCCAGGAGCTTT	TCAGGGCTCAGATTTGGGCATGGCAGA	696
Db	706	CAGCATCGGGGACA---TNTGCCCAGGAGNTTT	TCAGGGNTCAGATTTGGCATGCAGAGG	650
QY	697	AGAGGGCAGAGCGCTGGGGAGAAAGCCGCCAG	CACAGCCCCCTCGGAGGAGCATGT	756
Db	649	A---GCAGAGAGCCTTGGGGAG--AAGCCGGCC	AGACAGCCCCCTTNGAGAGGAGCATGT	594
QY	757	GACCTGCGTACAGAGCATCTTGGACGAATTCCT	TTC-AAACGTATGGCAGCTCATACCCC	815
Db	593	GACTTGGACACAGAGCATTTGGACGANTTCCT	TCAAACGTATGGCAGCTCATACCCCT	534
QY	816	TCAGCACTGATAGGCTAGTAGAAGCTGGAGACA	ATTTTCCAGCAGGAGTTTCCACCC	875
Db	533	TCAGCACTGATAGGCTAGTAGAAGTTGGAGGACA	ATTTTCCAGCAGGAGTTTCCACCC	474
QY	876	CTTCCAGGAGGGCCTGGTGTTCAGCTGATCCAG	TCTTACCAGCGGATGCCAGGCAATG	935
Db	473	TTTCCAGGAGGGCCTGGTGTTCAGCTNATCCAG	TCTTACCAGCGGATGCCAGGCAATG	414
QY	936	CCATGCTGAGGGCTTCCGAGTGGCTTATAAGCG	GCACGTGCTGACCATGGATGACTTG	995
Db	413	CCATGCTGAGGGCTTCCGAGTGGCTTATAAGCG	GCACGTGCTGACCATGGATGACTTG	354
QY	996	GGACCTTGATGACAGAACTGGCTCAATGACCA	GCGTGAACATGTATGGAGACCTGG	1055
Db	353	GGACCTTGATGACAGAACTGGCTCAATGACCA	GCGTGAACATGTATGGAGACCTGG	294
QY	1056	TCATGGACACAGTCCCTGAAAGGT	1080	
Db	293	TCATGGACACAGTCCCTGAAAGGT	269	

Search completed: December 16, 2002, 15:38:59  
Job time : 418 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: December 16, 2002, 13:18:49 ; Search time 64 Seconds  
(without alignments)  
9540.522 Million cell updates/sec

Title: US-09-848-852A-4  
Perfect score: 1991  
Sequence: 1 GCCCTCCTGTCCCGGCC.....TGCTCTGTCATAAAGATC 1991

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2.6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2.6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2.6/ptodata/2/ina/PTOTUS\_COMB.seq:\*  
6: /cgn2.6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70	3.5	51952	3	US-08-947-823-1
2	67.8	3.4	1890	6	Sequence 1, Appli Patent No. 5312912
3	67.2	3.4	51952	3	Sequence 1, Appli Patent No. 5312912
4	67	3.4	480	1	Sequence 4, Appli US-08-282-581-4
5	67	3.4	480	1	Sequence 4, Appli US-08-550-544-4
6	66.2	3.3	1890	6	Sequence 4, Appli Patent No. 5312912
7	65.8	3.3	480	1	Sequence 4, Appli US-08-282-581-4
8	65.8	3.3	480	1	Sequence 4, Appli US-08-550-544-4
9	65.6	3.3	297	3	Sequence 17, Appli US-08-616-368A-17
10	65.6	3.3	297	3	Sequence 17, Appli US-09-054-298-17
11	65.6	3.3	297	4	Sequence 17, Appli US-08-181-655-17
12	65	3.3	17656	4	Sequence 3, Appli US-09-433-579-3
13	64.6	3.2	128	1	Sequence 8, Appli US-07-922-723A-8
14	64.6	3.2	128	1	Sequence 8, Appli US-08-550-544-4
15	64.6	3.2	128	1	Sequence 8, Appli US-08-616-368A-17
16	64.6	3.2	128	1	Sequence 8, Appli US-08-074-275-8
17	64.6	3.2	128	2	Sequence 8, Appli US-08-480-366-8
18	64.6	3.2	297	1	Sequence 8, Appli US-07-952-277A-8
19	64.6	3.2	297	3	Sequence 17, Appli US-09-054-298-17
20	64.6	3.2	297	4	Sequence 17, Appli US-08-181-655-17
21	64.6	3.2	4254	2	Sequence 7, Appli US-08-443-639-7
22	63.8	3.2	1368	3	Sequence 5, Appli US-08-874-563-5
23	63.8	3.2	1368	3	Sequence 14, Appli US-08-577-483-14
24	63.8	3.2	4253	3	Sequence 7, Appli US-08-577-483-7
25	63.6	3.2	1368	3	Sequence 5, Appli US-08-874-563-5
26	63.6	3.2	1368	3	Sequence 14, Appli US-08-577-483-14
27	63.6	3.2	4253	3	Sequence 7, Appli US-08-577-483-7

28	63.4	3.2	128	1	US-07-922-723A-8	Sequence 8, Appli
29	63.4	3.2	128	1	US-07-799-828C-8	Sequence 8, Appli
30	63.4	3.2	128	1	US-08-074-275-8	Sequence 8, Appli
31	63.4	3.2	128	1	US-08-480-366-8	Sequence 8, Appli
32	63.4	3.2	128	2	US-07-952-277A-8	Sequence 8, Appli
33	62.8	3.2	87350	3	US-08-781-891-79	Sequence 79, Appli
34	62.8	3.2	87543	4	US-09-791-211-3	Sequence 3, Appli
35	62.2	3.1	4254	2	US-08-443-639-7	Sequence 7, Appli
36	62	3.1	15066	1	US-08-207-904-18	Sequence 18, Appli
37	61.4	3.1	15062	4	US-09-004-838-89	Sequence 89, Appli
38	61.4	3.1	87350	3	US-08-781-891-79	Sequence 79, Appli
39	61.4	3.1	87543	4	US-09-791-211-3	Sequence 3, Appli
40	61.4	3.1	246240	2	US-08-724-394A-20	Sequence 20, Appli
41	61.4	3.1	246240	2	US-08-724-394A-21	Sequence 21, Appli
42	61.4	3.1	246240	2	US-08-724-394A-22	Sequence 22, Appli
43	61.2	3.1	3072	4	US-09-522-217-55	Sequence 55, Appli
44	61.2	3.1	15062	4	US-09-004-838-89	Sequence 89, Appli
45	61	3.1	1906	1	US-08-207-904-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1  
US-08-947-823-1  
; Sequence 1, Application US/08947823  
; Patent No. 6114605  
; GENERAL INFORMATION:  
; APPLICANT: Williamson, Valerie M.  
; APPLICANT: Kaloshian, Isigouhi  
; APPLICANT: Yaghoobi, Jafar  
; APPLICANT: Bodeau, John  
; APPLICANT: Milligan, Stephen  
; TITLE OF INVENTION: Procedures and Materials for Conferring  
; TITLE OF INVENTION: Pest Resistance in Plants  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/947,823  
; FILING DATE: 09-OCT-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/18802  
; FILING DATE: 09-OCT-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/028,191  
; FILING DATE: 10-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-070210US  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 51952 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-947-823-1







```

; FILING DATE: 02-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18872
; REFERENCE/DOCKET NUMBER: 7956-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 162
; OTHER INFORMATION: /label= n
; OTHER INFORMATION: /note= "n-x=Unknown nucleotide"
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 356
; OTHER INFORMATION: /label= n
; OTHER INFORMATION: /note= "n-x=Unknown nucleotide"
US-08-550-544-4
Query Match 3.3%; Score 65.8; DB 1; Length 480;
Best Local Similarity 90.9%; Pred. No. 8e-08;
Matches 70; Conservative 0; Mismatches 7; Indels

QY 1884 CTTCAAACCTTTTATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA
DB 226 CTCTGCATTATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA
QY 1944 ATATATAAATATATAA 1960
||||| ||||| ||||| |
DB 286 ATATATATATATATA 302

RESULT 9
US-08-616-368A-17
; Sequence 17, Application US/08616368A
; Patent No. 5767262
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Haber, Edgar
; APPLICANT: Jain, Mukesh
; APPLICANT: Yet, Shaw-Fang
; TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/616,368A
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/022001
; TELECOMMUNICATION INFORMATION:

```

RESULT 12  
US-09-433-579-3/c  
; Sequence 3, Application US/09433579



Search completed: December 16, 2002, 14:21:22  
Job time : 161 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2002, 14:16:54 ; Search time 133 seconds  
(without alignments)  
5920.021 Million cell updates/sec

Title: US-09-848-852A-4  
Perfect score: 1991  
Sequence: 1 GCCTCCCTGTCGCCGACCC.....TGCTCTGGTCAATAAGATC 1991

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 35320 seqs, 197730502 residues  
Total number of hits satisfying chosen parameters: 710640

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1991	100.0	1991	10	US-09-848-852A-4
c 2	442	22.2	503	10	Sequence 1710, Ap
c 3	368	18.5	432	10	Sequence 835, App
c 4	92	4.6	435	10	Sequence 4735, Ap
c 5	84.8	4.3	106	10	Sequence 4719, Ap
c 6	72.6	3.6	217	10	Sequence 1226, Ap
c 7	70.8	3.6	217	10	Sequence 1226, Ap
c 8	70.6	3.5	138	10	Sequence 792, App
c 9	70.6	3.5	169	10	Sequence 741, App
c 10	70.6	3.5	170	10	Sequence 780, App
c 11	70.6	3.5	242	10	Sequence 781, App
c 12	70.4	3.5	123	10	Sequence 182, App
c 13	70.4	3.5	202	10	Sequence 5, Appli
c 14	70.4	3.5	229	10	Sequence 1068, Ap
c 15	70	3.5	170	10	Sequence 82, Appl
c 16	70	3.5	209	10	Sequence 1107, Ap
c 17	69.4	3.5	144	10	Sequence 38, Appl
c 18	69.4	3.5	741	10	Sequence 479, App
c 19	69.2	3.5	212	10	Sequence 1097, Ap

c 20	69	3.5	419	10	US-09-969-373-237
c 21	68.6	3.4	147	10	US-09-969-373-1517
c 22	68.6	3.4	252	10	US-09-969-373-1061
c 23	68.4	3.4	271	10	US-09-969-373-1313
c 24	68.4	3.4	313	10	US-09-969-373-1463
c 25	68	3.4	275	10	US-09-969-373-479
c 26	68	3.4	276	10	US-09-969-373-478
c 27	67.8	3.4	123	10	US-09-969-373-182
c 28	67.8	3.4	225	10	US-09-969-373-622
c 29	67.6	3.4	154	10	US-09-969-373-1052
c 30	67.6	3.4	170	10	US-09-969-373-82
c 31	67.6	3.4	230	10	US-09-969-373-110
c 32	67.6	3.4	230	10	US-09-969-373-111
c 33	67.6	3.4	275	10	US-09-969-373-479
c 34	67.6	3.4	276	10	US-09-969-373-478
c 35	67.6	3.4	419	10	US-09-969-373-237
c 36	67.4	3.4	147	10	US-09-969-373-1517
c 37	67.4	3.4	178	10	US-09-969-373-972
c 38	67.4	3.4	205	10	US-09-969-373-1270
c 39	67.4	3.4	223	10	US-09-969-373-1183
c 40	67.2	3.4	238	10	US-09-969-373-1498
c 41	67.2	3.4	271	10	US-09-969-373-1313
c 42	67.2	3.4	453	10	US-09-867-701-5706
c 43	67	3.4	138	10	US-09-969-373-792
c 44	67	3.4	229	10	US-09-969-373-1068
c 45	67	3.4	453	10	US-09-867-701-5706

ALIGNMENTS

RESULT 1  
US-09-848-852A-4  
; Sequence 4, Application US/09848852A  
; Patent No. US20020106373A1  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; Tang, Y. Tom  
; Corley, Neil C.  
; Guegler, Karl J.  
; Yue, Henry  
; Patterson, Chandra  
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH EGF SIGNALING  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/848,852A  
; FILING DATE: 07-May-2001  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/069,725  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0515 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1991 base pairs  
; TYPE: nucleic acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BEPINOT01
; CLONE: 2056178
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-848-852A-4

Query Match 100.0%; Score 1991; DB 10; Length 1991;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1991; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTCCCTGTC CCGGACCTCTTTTGATGCCCTCAGCAAGTGAAGAGAGAGAGAGAGG 60
DB 1 GCCTCCCTGTC CCGGACCTCTTTTGATGCCCTCAGCAAGTGAAGAGAGAGAGAGAGG 60

QY 61 AGGAGGAGGAGGATGAAGATGAAGAGAGAGGAAGTGGCAGCTTGGAGGCTGCCCCCAAGAT 120
DB 61 AGGAGGAGGAGGATGAAGATGAAGAGAGAGGAAGTGGCAGCTTGGAGGCTGCCCCCAAGAT 120

QY 121 GGAGTCAAGCTTGGAACTCCACAGCGGCCGCGCTTCCCGGCCACATCATCGAAAAACCT 180
DB 121 GGAGTCAAGCTTGGAACTCCACAGCGGCCGCGCTTCCCGGCCACATCATCGAAAAACCT 180

QY 181 GCTCACAGCGCGCGCGCGCAGCCATGAGAGCCCTTCCGGATGCTGCTACTCAAAAAAGCA 240
DB 181 GCTCACAGCGCGCGCGCGCAGCCATGAGAGCCCTTCCGGATGCTGCTACTCAAAAAAGCA 240

QY 241 CCTCGCTGACATTCACCTGGAAGCTTTGGGGGGCCACCGGGGGCGGGGGGGCGCTCG 300
DB 241 CCTCGCTGACATTCACCTGGAAGCTTTGGGGGGCCACCGGGGGCGGGGGGGCGCTCG 300

QY 301 CACACCCCAAGAACCATCTTTACCCAGCAAGGGGGTGGACGCCACAGTGCATCCC 360
DB 301 CACACCCCAAGAACCATCTTTACCCAGCAAGGGGGTGGACGCCACAGTGCATCCC 360

QY 361 CTGTGTTGCTGTTTGTACTCCCCCGGGGGCCACCTCCACCCCGGCTGGGTCTGCTAGGTG 420
DB 361 CTGTGTTGCTGTTTGTACTCCCCCGGGGGCCACCTCCACCCCGGCTGGGTCTGCTAGGTG 420

QY 421 CTCTCATGCTGAGGATGGGTGAGAGGGTCTCCACAGTGCCTCTGCGGGCCCCCATGG 480
DB 421 CTCTCATGCTGAGGATGGGTGAGAGGGTCTCCACAGTGCCTCTGCGGGCCCCCATGG 480

QY 481 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTGACCTGACTCGGGGCTCCTTT 540
DB 481 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTGACCTGACTCGGGGCTCCTTT 540

QY 541 CATGTACTCTGCCAACGGTTTGGGGGACAATCTGGGCCAGAGGGGAGCGCACTTGGC 600
DB 541 CATGTACTCTGCCAACGGTTTGGGGGACAATCTGGGCCAGAGGGGAGCGCACTTGGC 600

QY 601 ACCCCCTGATGCCAGCATCCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGCCCA 660
DB 601 ACCCCCTGATGCCAGCATCCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGCCCA 660

QY 661 GGAGCTTTTTCAGGGCTCAGATTGGGCATGGCAGAGAGGCAAGAGGCTTGGGGAGAA 720
DB 661 GGAGCTTTTTCAGGGCTCAGATTGGGCATGGCAGAGAGGCAAGAGGCTTGGGGAGAA 720

QY 721 AGCGGCCGACACAGCCCCCTTGGCAGAGGAGCATGTGACCTGCTACAGAGCATCTTGA 780
DB 721 AGCGGCCGACACAGCCCCCTTGGCAGAGGAGCATGTGACCTGCTACAGAGCATCTTGA 780

QY 781 CGAATTCCTTCAACGTATGGCAGCCTCATACCCCTCAGCACTGATGAGTAGAGAA 840
DB 781 CGAATTCCTTCAACGTATGGCAGCCTCATACCCCTCAGCACTGATGAGTAGAGAA 840

QY 841 GCTGGAGGACATTTCCAGCAGAGGTTTCCACCCCTTCCAGGAAGGGCCTGGTGTGCA 900
DB 841 GCTGGAGGACATTTCCAGCAGAGGTTTCCACCCCTTCCAGGAAGGGCCTGGTGTGCA 900

QY 901 GCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGCTGAGGGGCTTCCGAGTGGC 960

```

Db	901																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
----	-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--





```
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 4735
/ LENGTH: 435
/ TYPE: DNA
/ ORGANISM: Homo sapien
/ US-09-867-701-4735

Query Match          4.6%; Score 92; DB 10; Length 435;
Best Local Similarity 100.0%; Pred. No. 6.1e-14;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1427 CAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTTCAGCTTACCCAGGAGGACATGCC 1486
      |||||||
Db 344 CAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTTCAGCTTACCCAGGAGGACATGCC 403
      |||||||

QY 1487 AAACCTTCGTCGGCAGATCTACAGGAGCTGTG 1518
      |||||||
Db 404 AAACCTTCGTCGGCAGATCTACAGGAGCTGTG 435
      |||||||

RESULT 5
US-09-867-701-4719/c
/ Sequence 4719, Application US/09867701
/ Patent No. US20020132237A1
/ GENERAL INFORMATION:
/ APPLICANT: Jones, Robert
/ APPLICANT: Aglate, Paul A.
/ APPLICANT: Harlocker, Susan L.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ FILE REFERENCE: 210121.497
/ CURRENT APPLICATION NUMBER: US/09/867,701
/ NUMBER OF SEQ ID NOS: 10912
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 4719
/ LENGTH: 106
/ TYPE: DNA
/ ORGANISM: Homo sapien
/ US-09-867-701-4719

Query Match          4.3%; Score 84.8; DB 10; Length 106;
Best Local Similarity 97.7%; Pred. No. 1.8e-12;
Matches 86; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1901 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAA 1960
      |||||||
Db 106 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAA 47
      |||||||

QY 1961 TGCCACGGTCCTGCTCGTCAATAAAG 1988
      |||||||
Db 46 TGCCACGGTCCTGCTCGTCAATAAAG 19
      |||||||

RESULT 6
US-09-969-373-1226/c
/ Sequence 1226, Application US/09969373
/ Patent No. US20020133852A1
/ GENERAL INFORMATION:
/ APPLICANT: Effertz, Roger J.
/ APPLICANT: Hauge, Brian M.
/ TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
/ FILE REFERENCE: 38-10(52679)A
/ CURRENT APPLICATION NUMBER: US/09/969,373
/ CURRENT FILING DATE: 2001-10-02
/ PRIOR APPLICATION NUMBER: US 09/754,853
/ PRIOR FILING DATE: 2001-01-05
/ PRIOR APPLICATION NUMBER: US 09/760,427
/ PRIOR FILING DATE: 2001-01-13
/ PRIOR APPLICATION NUMBER: US 09/855,768
/ PRIOR FILING DATE: 2001-05-15
/ NUMBER OF SEQ ID NOS: 4593
/ SEQ ID NO 1226

/ LENGTH: 217
/ TYPE: DNA
/ ORGANISM: Glycine max
/ US-09-969-373-1226

Query Match          3.6%; Score 72.6; DB 10; Length 217;
Best Local Similarity 81.6%; Pred. No. 3.5e-09;
Matches 84; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1859 ACAATAGGATCACTGCCCTGCCAGATCTTCAAACTTTTATATATATATATATATATATA 1918
      |||||||
Db 180 ACTATATGTTAATTAATTTCCCCCATGATAATAATGTATATATATATATATATATATA 121
      |||||||
QY 1919 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAA 1961
      |||||||
Db 120 TATATATATATATATATATATATATATATATATATATATATATATATATATATATA 78
      |||||||

RESULT 7
US-09-969-373-1226
/ Sequence 1226, Application US/09969373
/ Patent No. US20020133852A1
/ GENERAL INFORMATION:
/ APPLICANT: Effertz, Roger J.
/ APPLICANT: Hauge, Brian M.
/ TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
/ FILE REFERENCE: 38-10(52679)A
/ CURRENT APPLICATION NUMBER: US/09/969,373
/ CURRENT FILING DATE: 2001-10-02
/ PRIOR APPLICATION NUMBER: US 09/754,853
/ PRIOR FILING DATE: 2001-01-05
/ PRIOR APPLICATION NUMBER: US 09/760,427
/ PRIOR FILING DATE: 2001-01-13
/ PRIOR APPLICATION NUMBER: US 09/855,768
/ PRIOR FILING DATE: 2001-05-15
/ NUMBER OF SEQ ID NOS: 4593
/ SEQ ID NO 1226
/ LENGTH: 217
/ TYPE: DNA
/ ORGANISM: Glycine max
/ US-09-969-373-1226

Query Match          3.6%; Score 70.8; DB 10; Length 217;
Best Local Similarity 91.5%; Pred. No. 1e-08;
Matches 75; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1880 AGATCTTCAAACTTTTATATATATATATATATATATATATATATATATATATATATATAT 1939
      |||||||
Db 58 AGAAGCTTCAACATATATATATATATATATATATATATATATATATATATATATATATATAT 117
      |||||||
QY 1940 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAA 1961
      |||||||
Db 118 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139
      |||||||

RESULT 8
US-09-969-373-792/c
/ Sequence 792, Application US/09969373
/ Patent No. US20020133852A1
/ GENERAL INFORMATION:
/ APPLICANT: Effertz, Roger J.
/ APPLICANT: Hauge, Brian M.
/ TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
/ FILE REFERENCE: 38-10(52679)A
/ CURRENT APPLICATION NUMBER: US/09/969,373
/ CURRENT FILING DATE: 2001-10-02
/ PRIOR APPLICATION NUMBER: US 09/754,853
/ PRIOR FILING DATE: 2001-01-05
/ PRIOR APPLICATION NUMBER: US 09/760,427
/ PRIOR FILING DATE: 2001-01-13
/ PRIOR APPLICATION NUMBER: US 09/855,768
/ PRIOR FILING DATE: 2001-05-15
/ NUMBER OF SEQ ID NOS: 4593
/ SEQ ID NO 792
```

```

; SEQ ID NO 792
; LENGTH: 138
; TYPE: DNA
; ORGANISM: G
US-09-969-373-79

```

[illegible]

## RESULT 9

```

US-09-969-373-741
; Sequence 741, Application US/09969373
; Patent No. US2002013852A1
; GENERAL INFORMATION:
; APPLICANT: Effertz, Roger J.
; APPLICANT: Hauge, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969, 373
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754, 853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760, 427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 741
; LENGTH: 169
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-741

```

[illegible]

RESULT 10

```

US-09-969-373-780
; Sequence 780, Application US/09969373
; Patent No. US2002013852A1
; GENERAL INFORMATION:
; APPLICANT: Effertz, Roger J.
; APPLICANT: Hauge, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15

```

```

; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 780
; LENGTH: 170
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-780

```

[illegible]

## RESULT 11

```

US-09-969-373-781
; Sequence 781, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
; APPLICANT: Effertz, Roger J.
; APPLICANT: Hauger, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; CURRENT FILING DATE: 2001-10-02
; PRIORITY APPLICATION NUMBER: US 09/754,853
; PRIORITY FILING DATE: 2001-01-05
; PRIORITY APPLICATION NUMBER: US 09/760,427
; PRIORITY FILING DATE: 2001-01-13
; PRIORITY APPLICATION NUMBER: US 09/855,768
; PRIORITY FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 781
; LENGTH: 242
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-781

```

[illegible]

## RESULT 12

```

US-09-969-373-182
;
; Sequence 182, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
; APPLICANT: Effertz, Roger J.
; APPLICANT: Hauge, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2002, 15:26:39 ; Search time 2556 Seconds

(without alignments)

12615.498 Million cell updates/sec

Title: US-09-848-852A-4

Perfect score: 1991

Sequence: 1 GCCTCCCTGTCCCGACCC.....TGCTCTGGTCAATAAGATC 1991

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estm:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_Other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	874	43.9	1071	14	BM806659
2	812.2	40.8	1134	14	BM804530
3	805.6	40.5	1060	14	BM806502
c	786.4	39.5	817	9	AL581080
5	748.8	37.6	1114	13	BM476787
6	735	36.9	867	14	BQ440555

7	720.4	36.2	1099	13	BM555097
8	716.6	36.0	905	14	BM898865
c	679.2	34.1	994	14	BQ057204
10	676.8	34.0	936	14	BQ433698
11	673.8	33.8	715	12	BG255499
c	634.2	31.9	662	9	AL579669
13	629.4	31.6	676	10	AW961869
14	628.4	31.6	804	12	BG870520
15	626	31.4	626	14	BM696661
16	617.8	31.0	985	13	BI456217
17	594.2	29.8	790	13	BI079603
c	591.4	29.7	626	9	AL578220
19	590.6	29.7	657	14	BM976678
c	589.2	29.6	646	14	BQ021355
21	576.6	29.0	728	12	BG865703
22	576	28.9	886	14	BQ942601
23	562.4	28.2	665	14	BQ186996
24	560.6	28.2	1076	12	BF164921
c	554	27.8	607	14	BQ001101
c	542.8	27.3	605	10	AW085552
c	539	27.1	835	13	BI732013
27	536	26.9	704	10	AV721960
28	532.6	26.8	620	10	AW578424
29	532.6	26.8	620	10	AW604359
30	523	26.3	552	10	AW964755
c	522.6	26.2	667	17	AZ399027
c	515.6	25.9	567	9	AI800802
c	509	25.6	689	12	BF608506
c	503.8	25.3	616	10	AW916790
c	503	25.3	1070	13	BI247525
37	498.2	25.0	643	14	N36589
c	493.4	24.8	529	12	BF195133
c	491	24.7	508	13	BM670002
c	485.6	24.4	599	13	BI017569
40	485.2	24.4	555	14	BQ220291
41	481.8	24.2	549	13	BM507574
42	478.6	24.0	819	12	BF139554
43	475.4	23.9	504	12	BF954315
44	474	23.8	632	12	BG083248
45					

ALIGNMENTS

RESULT 1	BM806659	BM806659	1071 bp	mRNA	linear	EST 05-MAR-2002
LOCUS	AGENCOURT_6542888	NIH_MGC_72	Homo sapiens	cDNA clone	IMAGE:5548958	
DEFINITION	5', mRNA sequence.					
ACCESSION	BM806659					
VERSION	BM806659.1	GI:19123482				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .					
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)					
JOURNAL	Unpublished (1999)					
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a> Tissue Procurement: ATCC/DCFD/PTP CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM12258 row: 1 column: 15 High quality sequence stop: 648. Location/Qualifiers					
FEATURES	1..1071					
source						

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5548958"
/clone_lib="NIH_MGC_72"
/tissue_type="melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT      255 a      264 c      271 g      278 t      3 others
ORIGIN
Query Match      43.9%; Score 874; DB 14; Length 1071;
Best Local Similarity 96.5%; Pred. No. 3.5e-154;
Matches 947; Conservative 0; Mismatches 26; Indels 8; Gaps 5;

QY 742 GCGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGGAGCAATCTCTCAACGATATGG 801
Db 1 GCGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGGAGCAATCTCTCAACGATATGG 60

QY 802 CAGCCTCATACCCCTCAGACATGATGAGTAGAGAGCTGAGAGCATTTTCCAGCA 861
Db 61 CAGCCTCATACCCCTCAGACATGATGAGTAGAGAGCTGAGAGCATTTTCCAGCA 120

QY 862 GGAGTTTTCACCCCTTCCAGGAAGGCGCTGGTGTTCAGCTGATCCAGTCTTACCAGCG 921
Db 121 GGAGTTTTCACCCCTTCCAGGAAGGCGCTGGTGTTCAGCTGATCCAGTCTTACCAGCG 180

QY 922 GATGCCAGGCAATGCATGTTGAGGGCTTCCAGATGGCTTATAAGCGGCACGTCTGAC 981
Db 181 GATGCCAGGCAATGCCATGTTGAGGGCTTCCAGATGGCTTATAAGCGGCACGTCTGAC 240

QY 982 CATGGATGACTTGGGACCTTGTATGACAGAACTGCTCAATGACAGCTGATGAACAT 1041
Db 241 CATGGATGACTTGGGACCTTGTATGACAGAACTGCTCAATGACAGCTGATGAACAT 300

QY 1042 GTATGAGACCTGGTGCATGACACAGTCCCTGAAAAGGTGCATTTCTCAATAGTTTCT 1101
Db 301 GTATGAGACCTGGTGCATGACACAGTCCCTGAAAAGGTGCATTTCTCAATAGTTTCT 360

QY 1102 CTATGATAACTCCGTACCAAGGGTTATGATGGGTGAAAAGTGGACCAAAACGTGGA 1161
Db 361 CTATGATAACTCCGTACCAAGGGTTATGATGGGTGAAAAGTGGACCAAAACGTGGA 420

QY 1162 CATCTTCAATAAGAGCTACTGCTAATCCCATCCACCTGGAGTGCATTTGGTCCCTCAT 1221
Db 421 CATCTTCAATAAGAGCTACTGCTAATCCCATCCACCTGGAGTGCATTTGGTCCCTCAT 480

QY 1222 CTCTGTTGATGTGAGCGGACGACCATCATCTATTTTGTACTCGACGCTACCTTAACCG 1281
Db 481 CTCTGTTGATGTGAGCGGACGACCATCATCTATTTTGTACTCGACGCTACCTTAACCG 540

QY 1282 CCCTGCTCCCTAAGCATATTTGCCAAGTATCTACAGGAGAGGGGTAAAGAACCCGACT 1341
Db 541 CCCTGCTCCCTAAGCATATTTGCCAAGTATCTACAGGAGAGGGGTAAAGAACCCGACT 600

QY 1342 GGATTTCCACAGGGCTGGAAGGTTACTTCAAAATGAATGTGCCAGCAGCAATAATGA 1401
Db 601 GGATTTCCACAGGGCTGGAAGGTTACTTCAAAATGAATGTGCCAGCAGCAATAATGA 660

QY 1402 CAGTGACTGTGGTGCTTTTGTGTGAGTACTGCAAGCATCTTGGCCCTGTCTCAGCCATT 1461
Db 661 CAGTGACTGTGGTGCTTTTGTGTGAGTACTGCAAGCATCTTGGCCCTGTCTCAGCCATT 720

QY 1462 CAGCTTCAACCCAGCAGACATGCCAAACTTCGTGGCAGATCTACAAGAGCTGTGTCA 1521
Db 721 CAGCTTCAACCCAGCAGACATGCCAAACTTCGTGGCAGATCTACAAGAGCTGTGTCA 780

QY 1522 CT-GCAAACTCACTGTGTGAGCTCTGTACCCAGACCCCAAGGCCCAATAATGGGAAGGA 1580
Db 781 CTGGCAAACTCACTGTGTGAGCTCTGTACCCAGACCCC-AGCCCAATAATGGGAAGGA 839

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5554637"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; Oligo-dt primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT      261 a      326 c      311 g      236 t
ORIGIN
Query Match      40.8%; Score 812.2; DB 14; Length 1134;
Best Local Similarity 96.1%; Pred. No. 1.4e-142;
Matches 854; Conservative 0; Mismatches 33; Indels 2; Gaps 2;

QY 286 GCGCGGGGGCTCGCACACACCCCAAGAACATCTTTTACCCCAAGAGGGGTGCGACGC 345
Db 1 GCGCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 60

QY 346 CACAGGTGCGATCCCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 405
Db 61 CACAGGTGCGATCCCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 120

QY 406 TGGGCTCTGTAGTGTCTCTCATGGTGTGAGGTGGGTGAGAGGTCTCCACCAAGTCCCT 465
Db 121 TGGGCTCTGTAGTGTCTCTCATGGTGTGAGGTGGGTGAGAGGTCTCCACCAAGTCCCT 180

QY 466 CTGGGCCCCCCCCTGAGGAGAGATGGACTCAGTGGACTCCAAAGTCTCTCTGGACCCCTG 525
```

```
Db 181 CTGGGCCCCATGAGGAAGATGACTCAGGTGACTCCAAAGTCTCTCTGGACCCTG 240
QY 526 ACTCGGGCTCTCTTCATGTACTCTGCCCAACGGTTTGGGGACAATCTGGCCAGAG 585
Db 241 ACTCGGGCTCTCTTCATGTACTCTGCCCAACGGTTTGGGGACAATCTGGCCAGAG 300
QY 586 GGGAGCGCA-CCTTGACCCCTGTATGCGCAGCATCTCATAGCAATGTGTGCACATCG 644
Db 301 GGGAGCGCAGCTTGGCACCCCTGTATGCGCAGCATCTCATAGCAATGTGTGCACATCG 360
QY 645 GGGACCATGTGGCCAGAGCTTTTCAGGGCTCAGATTGGGCAATGGGAGAGGAG 704
Db 361 GGGACCATGTGGCCAGAGCTTTTCAGGGCTCAGATTGGGCAATGGGAGAGGAG 420
QY 705 AGAGGCTGGGAGAGAAAGCGCCAGCACAGCCCTCGGAGAGGAGCATGTGACTGCG 764
Db 421 AGAGGCTGGGAGAGAAAGCGCCAGCACAGCCCTCGGAGAGGAGCATGTGACTGCG 480
QY 765 TACAGAGCATCTTGACGAATTCCTTCAAAAGTATGGCAGCCTCATACCCCTCAGCACTG 824
Db 481 TACAGAGCATCTTGACGAATTCCTTCAAAAGTATGGCAGCCTCATACCCCTCAGCACTG 540
QY 825 ATGAGGTATGAGAAAGCTGGAGGACATTTCCAGCAGAGGTTTCCACCCCTTCCAGGA 884
Db 541 ATGAGGTATGAGAAAGCTGGAGGACATTTCCAGCAGAGGTTTCCACCCCTTCCAGGA 600
QY 885 AGGGCTGGTGTGAGAGCTGATCCAGCTTACCAGCGATGCCAGGCAATGCCATGGTGA 944
Db 601 AGGGCTGGTGTGAGAGCTGATCCAGCTTACCAGCGATGCCAGGCAATGCCATGGTGA 660
QY 945 GGGGCTCCGAGTGTATTAAGCGGCACGCTGACCATGGATGACTTGGGAGCTTGT 1004
Db 661 GGGGCTCCGAGTGTATTAAGCGGCACGCTGACCATGGATGACTTGGGAGCTTGT 720
QY 1005 ATGACAGAACTGGCTCAATGACAGGTGATGAACATGTATGAGACCTGGTCAATGGACA 1064
Db 721 ATGACAGAACTGGCTCAATGACAGGTGATGAACATGTATGAGACCTGGTCAATGGACA 780
QY 1065 CAGTCCCTGAAAAGTGTATTTCTCAATAGTTTCTTATGATATAAATCCGTACCAAG 1124
Db 781 CA-TCCCTGAAAAGTGTATTTCTCAATAGTTTCTTATGATATAAATCCGTACCAAG 839
QY 1125 GTTATGATGGGTGAAAAGTGGACAAAACGCTGGACATCTTCAATAA 1173
Db 840 GTTATGAGGGGTGAAAAGTGGACAAAACGCTGGACATCTTCAATAA 888
```

```
RESULT 3
BM806502
LOCUS BM806502
DEFINITION AGENCOURT_6542932 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5548865
5', mRNA sequence.
ACCESSION BM806502
VERSION BM806502.1 GI:19123325
KEYWORDS EST.
SOURCE human.
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Distribution by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

```
Plate: LLAM12258 row: h column: 18
High quality sequence stop: 652.
FEATURES
    source
        1..1060
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:5548865"
        /clone_lib="NIH_MGC_72"
        /tissue_type="melanotic melanoma"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
        Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
        Average insert size 2 kb. Library constructed by Life
        Technologies."
BASE COUNT 248 a 272 c 304 g 236 t
ORIGIN
```

```
Query Match 40.5%; Score 805.6; DB 14; Length 1060;
Best Local Similarity 95.8%; Pred. No. 2.4e-141;
Matches 883; Conservative 0; Mismatches 29; Indels 10; Gaps 5;
QY 503 TCCAAAGTCTCTCTGGACCCCTGACTCGGGCTCCTTTTCATCTCTCTGCCCAACGGTTT 562
Db 1 TCCAAAGTCTCTCTGGACCCCTGACTCGGGCTCCTTTTCATCTCTCTGCCCAACGGTTT 60
QY 563 TGGGGACAATCTGGCCAGAAAGGGAGGCGCA-CCTGGCACCCCTGTATGCCAGCATCCT 621
Db 61 TGGGGACAATCTGGCCAGAAAGGGAGGCGCAGCTTGGCACCCCTGTATGCCAGCATCCT 120
QY 622 CATCAGCAATGTGTGAGCATCGGGACCATGTGCCCCAGGAGCTTTTTCAGGGGTCTAGA 681
Db 121 CATCAGCAATGTGTGAGCATCGGGACCATGTGCCCCAGGAGCTTTTTCAGGGGTCTAGA 180
QY 682 TTTGGGCATGGCAGAAAGGAGAGAGGCTTGGGAGAAAGCGGCCACACAGACCCCT 741
Db 181 TTTGGGCATGGCAGAAAGGAGAGAGGCTTGGGAGAAAGCGGCCACACAGACCCCT 240
QY 742 GCGAGAGAGCATGTGACCTGCTACAGAGCATCTTGGAGCAATCTCTTCAACAGTATGG 801
Db 241 GCGAGAGAGCATGTGACCTGCTACAGAGCATCTTGGAGCAATCTCTTCAACAGTATGG 300
QY 802 CAGCCTCATACCCCTCAGCACTGTAGGTAGTAGAGAGCTGGAGGACATTTTCCAGCA 861
Db 301 CAGCCTCATACCCCTCAGCACTGTAGGTAGTAGAGAGCTGGAGGACATTTTCCAGCA 360
QY 862 GGAGTTTCCACCCCTTCCAGGAAGGCTGTGTGAGCTGATCCAGCTCTTACCAGCG 921
Db 361 GGAGTTTCCACCCCTTCCAGGAAGGCTGTGTGAGCTGATCCAGCTCTTACCAGCG 420
QY 922 GATGCCAGCAATGCCATGGTGAGGGCTTCCGAGTGGCTTATAAGCGGCAGCTGCTGAC 981
Db 421 GATGCCAGCAATGCCATGGTGAGGGCTTCCGAGTGGCTTATAAGCGGCAGCTGCTGAC 480
QY 982 CATGATGACTTTGGGGACCTTGTATGGACAGAACTGGCTCAATGACCAAGTATGAACAT 1041
Db 481 CATGATGACTTTGGGGACCTTGTATGGACAGAACTGGCTCAATGACCAAGTATGAACAT 540
QY 1042 GTATGAGACCTTGGTCAATGACAGCATGCTCCTGAAAAGGTGCAATTTCTCAATAGTTCTT 1101
Db 541 GTATGAGACCTTGGTCAATGACAGCATGCTCCTGAAAAGGTGCAATTTCTCAATAGTTCTT 600
QY 1102 CTATGATAAATCCGTTACCAAGGTTATGATGGGTGAAAAGGTGACCAAAACGTGGA 1161
Db 601 CTATGATAAATCCGTTACCAAGGTTATGATGGGTGAAAAGGTGACCAAAACGTGGA 660
QY 1162 CATCTTCAATAAGAGCTACTGCTAATCCCATCCACCTGGAGGTCATTGGTCCCTCAT 1221
Db 661 CATCTTCAATAAGAGCTACTGCTAAT -CCCATCCACCTGGAGGTGTCATTGGTCCCTCAT 719
QY 1222 CTCCTGTTGATGAGGCGAGCAGCACCATCACCTATTTTGTACTGCGAGGTCACCTTAACCG 1281
Db 720 CTCCTGTTGATGAGGCGAGCAGCACCATCACCTATTTTGTACTGCGAGGTCACCTTAACCG 779
```





```
/clone="IMAGE:5555330"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."
BASE COUNT      250 a   315 c   363 g   186 t
ORIGIN
Query Match      37.6%; Score 748.8; DB 13; Length 1114;
Best Local Similarity 96.6%; Pred. No. 1.1e-120;
Matches 840; Conservative 0; Mismatches 22; Indels 8; Gaps 7;
QY 29 TGCCTCAGCAAGTGAAGAGGAGGAAGAGAGGAGGAGGAGGAGGATGAAGATGAAGAGGA 88
Db 1 TGCCTCAGCAAGTGAAGAGGAGAGAGAGGAGGAGGAGGAGGATGAAGATGAATAGGA 60
QY 89 GGAAGTGGAGCTTGGAGGCTGCCCCCAAGATGGAGTGCAGCTGGGAACCTCCCGCGGCC 148
Db 61 GGAAGTGGAGCTTGTAGGCTGCCCCCAAGATGGAGTGCAGCTGGGAACCTCCCGCGGCC 120
QY 149 CGGCCCTCCCGCCCACTCATCGAAACCTGCTCACAGCCCGCCCGCGAGCCATGAG 208
Db 121 CGGCCCTCCCGCCCACTCATCGAAACCTGCTCACAGCCCGCCCGCGAGCCATGAG 180
QY 209 AGCCTTCGGGATGCTGCTTACTCAAAAAGCACCTCGCTGACATTCCTGGAAGCTTGG 268
Db 181 AGCCTTCGGGATGCTGCTTACTCAAAAAGCACCTCGCTGACATTCCTGGAAGCTTGG 240
QY 269 GGGGGCCACACCGGGCGCGCGGGGCTCGCACACCCCAAGAACCATCTTTTACCCCA 328
Db 241 GGGGGCCACACCGGGCGCGCGGGGCTCGCACACCCCAAGAACCATCTTTTACCCCA 300
QY 329 GCAAGGGGTGCGAGCCACAGGTGCCATCCCGCTGTGTGCTTTTGACTCCCCCGGGG 388
Db 301 GCAAGGGGTGCGAGCCACAGGTGCCATCCCGCTGTGTGCTTTTGACTCCCCCGGGG 360
QY 389 GCCAGCTCCACCCCGCTGGGTGCTGCTAGGTGCTCTCATGGCTGAGGATGGGGTGAGAG 448
Db 361 GCCAGCTCCACCCCGCTGGGTGCTGCTAGGTGCTCTCATGGCTGAGGATGGGGTGAGAG 420
QY 449 GTCTCACACAGTGCCTCTGGGCCCGCCCATGAGGAAAGATGGACTCAGTGACTCCAAA 508
Db 421 GTCTCACACAGTGCCTCTGGGCCCGCCCATGAGGAAAGATGGACTCAGTGACTCCAAA 480
QY 509 GTCTCTCTGGACCTGACTCGGGCTCTCTTCATGTACTCTGCCCAACGGTTTGGGGG 568
Db 481 GTCTCTCTGGACCTGACTCGGGCTCTCTTCATGTACTCTGCCCAACGGTTTGGGGG 540
QY 569 ACAATCTGGGCAGAGGGGAGCGCA-CTTGGCACCCCTGTATGCGAGCATCTCATCAG 627
Db 541 ACAATCTGGGCAGAGGGGAGCGCAGCTTGGCACCCCTGTATGCGAGCATCTCATCAG 600
QY 628 CAATGTGTGACATCGGGGACCATGTGGCCAGGAGCTTTTTCAGGGGCTCAGATTTGGG 687
Db 601 CAATGTGTGACATCGGGGACCATGTGGCCAGGAGCTTTTTCAGGGGCTCAGATTTGGG 660
QY 688 CATGCGAGAGGAGGAGAGGAGGCTTGGGGAGAAAGCGCCAGCAG-CCCGCTGGGA 745
Db 661 CATGCGAGAGGAGGAGAGGAGGCTTGGGGAGAAAGCGCCAGCAGCAGCCCGCTGGGA 720
QY 746 GAGGAGCATGTGACCTGCGCTACAGAGCAT-CTTGAGCAAAAT-CTTCAAAAGTATGCA 803
Db 721 GAAGAGCATGTGACCTGCGCTACAGAGGATCTTGGACCAATTCCTTCAAAAGGATGGA 780
QY 804 GCCTCATACCCCTCAGCACTGTATGAGGTAGTA--GAAGCTGGAGGACATTTTCC-AGC 860
Db 781 GCCTCATACCCCTCAGCACTGTATGAGGTAGTA--GAAGCTGGAGGACATTTTCCAGA 840
QY 861 AGGAGTTTTCACCCCTTCCAGGAAGGGCC 890
Db 841 AGAAGTTTTCACCCCTTCCAGGAAGGGCC 870
```

```
RESULT 6
BO440555
LOCUS      BO440555
DEFINITION AGENCOURT_7827074 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6153067
5', mRNA sequence.
ACCESSION BO440555
VERSION    BO440555
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 867)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM13491 row: o column: 20
          High quality sequence stop: 645.
          Location/Qualifiers
            1..867
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:6153067"
              /clone_lib="NIH_MGC_67"
              /tissue_type="retinoblastoma"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
              Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
              Average insert size 1.75 kb. Library constructed by Life
              Technologies."
BASE COUNT 191 a   266 c   261 g   148 t   1 others
ORIGIN
Query Match      36.9%; Score 735; DB 14; Length 867;
Best Local Similarity 98.2%; Pred. No. 4.5e-128;
Matches 785; Conservative 0; Mismatches 10; Indels 4; Gaps 4;
QY 30 GCCTCAGCAAGTGAAGAGGAGGAAGAAGAGGAGGAGGAGGATGAAGATGAAGAGGAG 89
Db 12 GCCTCAGCAAGTGAAGAGGAGGAAGAAGAGGAGGAGGAGGATGAAGATGAAGAGGAG 71
QY 90 GAAGTGGAGCTTGGAGGCTGCCCCCAAGATGGATGAGTGCAGTGGGAACCTCCCGAGGGCCC 149
Db 72 GAAGTGGAGCTTGGAGGCTGCCCCCAAGATGGATGAGTGCAGTGGGAACCTCCCGAGGGCCC 131
QY 150 CGCCCTTCGCCCGCCCACTCATCGAAAAACCTGCTCACAGCGCCGCCGCGAGGCCATGAGA 209
Db 132 CGCCCTTCGCCCGCCCACTCATCGAAAAACCTGCTCACAGCGCCGCCGCGAGGCCATGAGA 191
QY 210 GCCTTCGGGATGCTCTTACTCAAAAAGCACCTCGCTGACATTCCTACTGGGAAGCTTTGG 269
Db 192 GCCTTCGGGATGCTCTTACTCAAAAAGCACCTCGCTGACATTCCTACTGGGAAGCTTTGG 251
QY 270 GGGCGCCACCGGGCGCGGGGGCGCTCGCACACCCCAAGAACCATCTTTTACCCCGAG 329
Db 252 GGGCGCCACCGGGCGCGGGGGCGCTCGCACACCCCAAGAACCATCTTTTACCCCGAG 311
QY 330 CAAGGGGTGCGACCGCCACAGGTGCCATCCCGCTGTGTCGTTTGTACTCCCGCGGGG 389
Db 312 CAAGGGGTGCGACCGCCACAGGTGCCATCCCGCTGTGTCGTTTGTACTCCCGCGGGG 371
QY 390 CCACCTCCACCCCGCTGGGTGCTGTAGTGCTCTCATGGTGAAGGATGGGTGAGAGGG 449
```



TITLE  
JOURNAL  
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LRAM13564 row: o column: 14  
High quality sequence stop: 614.

FEATURES  
Source

1. 905  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6181093"  
/clone\_lib="Lupski\_dorsal\_root\_ganglion"  
/sex="male"  
/tissue\_type="dorsal root ganglia"  
/dev\_stage="adult, 36 yr"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1:  
Noti; Site\_2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCCACGGCTCCG-3' and  
5'-GACTAGTCTAGATCGCAGCGCCGCT(15)-3'. Size selected >  
1 kb for average insert length 1.7 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine) and is available through Life  
Technologies."

BASE COUNT 212 a 238 c 220 g 235 t  
ORIGIN

Query Match 36.0%; Score 716.6; DB 14; Length 905;  
Best Local Similarity 96.1%; Pred. No. 1.3e-124;  
Matches 756; Conservative 0; Mismatches 29; Indels 2; Gaps 2;

QY 946 GGGCTTCGAGTGGTTTATAAGCGGCACGTGCTGACCATGGATGGTGGGACCTGTGA 1005\*  
|||||  
Db 1 GGGCTTCGAGTGGTTTATAAGCGGCACGTGCTGACCATGGATGGTGGGACCTGTGA 60  
QY 1006 TGSACAGACTGGCTCAATGACAGGTGATGACATGATGAGACCTGGTCAATGACAC 1065  
|||||  
Db 61 TGSACAGACTGGCTCAATGACAGGTGATGACATGATGAGACCTGGTCAATGACAC 120  
QY 1066 AGTCCTGAAAAAGGTGCATTTCTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGG 1125  
|||||  
Db 121 AGTCCTGAAAAAGGTGCATTTCTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGG 180  
QY 1126 TTATGATGGGTGAAAAGTGGACCAAAACCTGGACATCTTCAATAAGAGGACTACTGCT 1185  
|||||  
Db 181 TTATGATGGGTGAAAAGTGGACCAAAACCTGGACATCTTCAATAAGAGGACTACTGCT 240  
QY 1186 ATATCCCATCCACCTGGAGTGCATTTGTCCTCATCTCTGTTGATGTGAGCGGACGAC 1245  
|||||  
Db 241 ATATCCCATCCACCTGGAGTGCATTTGTCCTCATCTCTGTTGATGTGAGCGGACGAC 300  
QY 1246 CATCACCTATTTTGACTCGCAGCGTACCCTAAACCGCGCTGCCCTTAAGCATATTGCCAA 1305  
Db 301 CATCACCTATTTTGACTCGCAGCGTACCCTAAACCGCGCTGCCCTTAAGCATATTGCCAA 360  
QY 1306 GTATCTACAGGCAGGCGGTAAAGAAACCGACTGGATTTCCACCGGGCTGAAAGG 1365  
Db 361 GTATCTACAGGCAGGCGGTAAAGAAACCGACTGGATTTCCACCGGGCTGAAAGG 420  
QY 1366 TTACTTCAAAATGAATGGCCAGGCAGCAATTAATGACAGTGAAGTGGTCTTTTGTGTT 1425  
Db 421 TTACTTCAAAATGAATGGCCAGGCAGCAATTAATGACAGTGAAGTGGTCTTTTGTGTT 480  
QY 1426 GCAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTACAGCTTCACCCAGCAGCATGCC 1485

Db 481 GCAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTACGCTTCACCCAGCAGCATGCC 1540  
QY 1486 CAAACTTCGTCGGCAGATCTACAAGAGCTGTGTCACTCAAACTCACTGTGTGAGCCTC 1545  
Db 541 CAAACTTCGTCGGCAGATCTACAAGAGCTGTGTCACTCAAACTCACTGTGTGAGCCTC 600  
QY 1546 GTACCCCCAGA-CCCCAAGCCCAATAATGGGAAGGAGACATGGAGTCCCTTCCCAAGAA 1604  
|||||  
Db 601 GTACCCCCAGACCCCAAGCCCAATAATGGGAAGGAGACATGGAGTCCCTTCCCAAGAA 560  
QY 1605 ACTCCAGTTCCTTCTCTCTCTTTCCTCTTCCCACTCACTT-CCCTTTGGTTTTTCATATT 1663  
|||||  
Db 661 ACTCCAGTTCCTTCTCTCTCTTTCCTCTTCCCACTCACTTCCCTTTGGTTTTTCATATT 720  
QY 1664 TAAATGTTTCAATTTCTGTATTTTTTTTCTTTGAGAGAACTACTGTGATTTCTGATGT 1723  
Db 721 TAAATGTTTCAATTTCTGTATTTTTTTTCTTTGAGAGAACTACTGTGATTTCTGATGT 780  
QY 1724 GCAGGGG 1730  
Db 781 GAATGGG 787

RESULT 9  
LOCUS BQ057204  
DEFINITION AGENCOURT\_6769328 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5812399  
5', mRNA sequence.  
ACCESSION BQ057204  
VERSION BQ057204.1 GI:19816544  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 994)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Lou Staudt  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCOM2062 row: m column: 08  
High quality sequence stop: 620.

FEATURES  
source  
Location/Qualifiers  
1. 994  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5812399"  
/clone\_lib="NIH\_MGC\_99"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 278 a 228 c 232 g 256 t  
ORIGIN

Query Match 34.1%; Score 679.2; DB 14; Length 994;  
Best Local Similarity 95.0%; Pred. No. 1.3e-117;

Matches		720;	Conservative	0;	Mismatches	18;	Indels	20;	Gaps	1;
Qy	1231	TGTGAGCGCAGCCACCATACCTATTTTGTGCTGGCAGCGTACCTAAACCGCGCTGCC	1290							
Db	795	TGAGGGGAGCGCCCGCTCTTATTTGACTCGCAGGGTCCCTTAACCGCGCTCCC	736							
Qy	1291	TAGCATATTGCCAAGTATCTACAGGCAGAGGGGTAAAGAAACCGACTGATTTCCA	1350							
Db	735	TAGCATATTGCCAAGTATCTACAGGCAGAGGGGTAAAGAAACCGACTGATTTCCC	676							
Qy	1351	CCAGGGCTGGAAGGTTTACTTCAAAATGAATGTGGCCAGGCAGAAATACAGTGA	1410							
Db	675	CCAGGGCTGGAAGGTTTACTTCAAAATGAATGTGGCCAGGCAGAAATACAGTGA	616							
Qy	1411	TGTTGCTTTTGTGTGAGTACTGCAAGCATCTGGCCCTCTCTCAGCCATTCAGCTTCA	1470							
Db	615	TGTTGCTTTTGTGTGAGTACTGCAAGCATCTGGCCCTCTCTCAGCCATTCAGCTTCA	556							
Qy	1471	CCAGCAGGACATGCCAAACTTCGTGGCGAGATCTACAAGAGCTGTGTCACTGCAAACT	1530							
Db	555	CCAGCAGGACATGCCAAACTTCGTGGCGAGATCTACAAGAGCTGTGTCACTGCAAACT	496							
Qy	1531	CACGTGTGAGCTCTGTACCCAGACCCCAAGCCCATAAATGGAGGGAGACATGGGAG	1590							
Db	495	CACGTGTGAGCTCTGTACCCAGACCCCAAGCCCATAAATGGAGGGAGACATGGGAG	436							
Qy	1591	TCCTCTCCCAAGAACTCCAGTTTCCCTTCTCTGCTCTTCCCTTCCCACTCACCTTCC	1650							
Db	435	TCCTCTCCCAAGAACTCCAGTTTCCCTTCTCTGCTCTTCCCTTCCCACTCACCTTCC	376							
Qy	1651	GGTTTTTCATATTAATGTTTCAATTTCTGTATTTTCTTTTCTTGAGAGAACTTGT	1710							
Db	375	GGTTTTTCATATTAATGTTTCAATTTCTGTATTTTCTTTTCTTGAGAGAACTTGT	316							
Qy	1711	TGATTTCTGATGTCAGGGGGTGGCTACAGAAAGCCCTTCTCTCTGTTTGAGGG	1770							
Db	315	TGATTTTGTGATGTCAGGGGGTGGCTACAGAAAGCCCTTCTCTCTGTTTGAGGG	256							
Qy	1771	GAGTGTGGCCCTGTGGCTGGTGGAGCAGTATCTTCCCTTCCCGTGCAGGAGCA	1830							
Db	255	GAGTGTGGCCCTGTGGCTGGTGGAGCAGTATCTTCCCTTCCCGTGCAGGAGCA	196							
Qy	1831	GGAATCAGTGTGGGGGTGGGGGGGACAAATAGGATCACTCCCTGCCAGATCTTCAA	1890							
Db	195	GGAATCAGTGTGGGGGTGGGGGGGACAAATAGGATCACTCCCTGCCAGATCTTCAA	136							
Qy	1891	CTTTTA	1950							
Db	135	CTTT-----TA	96							
Qy	1951	AATATATAATGCACGGTCTCTCTGGTCAATAAG	1988							
Db	95	AATATATAATGCACGGTCTCTCTGGTCAATAAG	58							

RESULT 10  
LOCUS BQ433698  
DEFINITION AGENCOURT\_7770396 NIH\_MGC\_70 Homo sapiens cDNA clone IMAGE:6025493  
5', mRNA sequence.  
ACCESSION BQ433698  
VERSION BQ433698.1 GI:21172774  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 936)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)

Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM13237 row: 1 column: 06  
High quality sequence start: 41  
High quality sequence stop: 300.  
Location/Qualifiers  
I. .936  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6025493"  
/clone\_lib="NIH\_MGC\_70"  
/tissue\_type="epithelioid carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."  
BASE COUNT 200 a 296 c 283 g 157 t  
ORIGIN

Query Match 34.0%; Score 676.8; DB 14; Length 936;  
Best Local Similarity 96.0%; Pred. No. 3.8e-117;  
Matches 726; Conservative 0; Mismatches 27; Indels 3; Gaps 3;

Qy 1 GCCTCCTCTGCCCGACCCCTCTTTTGTATGCCTCAGCAAGTGAAGAGGAGGAGAGAGG 60  
Db 87 GCCTCCTCTGCCCGACCCCTCTTTTGTATGCCTCAGCAAGTGAAGAGGAGGAGAGAGG 146  
Qy 61 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGTGGCAGCTTGGAGGCTGCCCGCAAGAT 120  
Db 147 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGTGGCAGCTTGGAGGCTGCCCGCAAGAT 206  
Qy 121 GGAGTCAGCTGGGAACCTCCAGCGGCCCGCCCTTCCCGCCCTCATCGAAAAACCT 180  
Db 207 GGAGTCAGCTGGGAACCTCCAGCGGCCCGCCCTTCCCGCCCTCATCGAAAAACCT 266  
Qy 181 GGTACAGCGCCGCCCGCGAGGCATGAGAGCCTTCC - GGATGCTGCTCTACTCAAAAAGC 239  
Db 267 GGTACAGCGCCGCCCGCGAGCGCGATAGCGCTCTGTATGCTGCTCTACTCAAAAAGC 326  
Qy 240 ACCTCCTCAGATTCACATGGAGCTTTGGGGCGCCACCGGGCGCGGGCGGGGGCCCTC 299  
Db 327 ACCTCCTCAGATTCACATGGAGCATTTGGGGCGCCACCGGGCGCGGGGGGCCCTC 386  
Qy 300 GCACACCCCAAGAACCATCTTTCACCCCAAGAGGGGTGCCAGCCACAGGTCGCCATCC 359  
Db 387 GCACACCCCAAGAACCATCTTTCACCCCAAGAGGGGTGCCAGCCACAGGTCGCCATCC 446  
Qy 360 CCTGTGTTGCTTTGACTTCCCGCGGGGCCACCTCCACCGCGGTGGGTCTGTAGGT 419  
Db 447 CCTGTGTTGCTTTGACTTCCCGCGGGGCCACCTCCACCGCGGTGGGTCTGTAGGT 506  
Qy 420 GCTCTCATGCTGAGGATGGGTGAGAGGCTTCCACACAGTGCCTCTGGGGCCCCCATG 479  
Db 507 GCTCTCATGCTGAGGATGGGTGAGAGGCTTCCACACAGTGCCTCTGGGGCCCCCATG 566  
Qy 480 GAGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTTGAGCCCTGACTCGGGCCCTT 539  
Db 567 GAGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTTGAGCCCTGACTCGGGCCCTT 626  
Qy 540 TCATGTACTCTGCCAACCGTTTGGGGGACAAATCTGGGCCAGAGGGAGGAGGCGCA -CTTG 598  
Db 627 TCATGTACTCTGCCAACCGTTTGGGGGACAAATCTGGGCCAGAGGGAGGAGGCGCACTTG 686  
Qy 599 GCACCCCTGATGCCAGCATCTCATCAGCAATGTCTGAGCATCGGGACCATGTGGCC 658  
Db 687 GCACCCCTGATGCCAGCATCTCATCAGCAATGTGGGGGAGCATCGGGACCATGTGGCC 746

FEATURES  
source

```

QY 659 CAGGAGCTTTTTCAGGCTCAGATTTGGCATGGCAGAGAGGACAGAGAGGCTT-GGGGA 717
|||||
Db 747 CAGGAGCTTTTTCAGGCTCAGATTTGGCATGGCAGAGAGGACACAAAGGCTTGGGGGA 806
|||||
QY 718 GAAAGCGGCCAGCAGAGCCCTTGGGAGAGAGCA 753
|||||
Db 807 AAAACCGGCCAGCAACAGCCCTTTCGAAGAGGA 842
|||||

RESULT 11
BG255499
LOCUS 602367640F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4476128 5',
DEFINITION mRNA sequence.
ACCESSION BG255499
VERSION BG255499.1 GI:12765237
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS 1 (bases 1 to 715)
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgi.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DMP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM0302 row: o column: 09
High quality sequence stop: 710.
Location/Qualifiers
1..715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4476128"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
Note="Organ: prostate; Vector: pCMV-SPORT6; Site:1; NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 164 a 188 c 175 g 188 t
ORIGIN
Query Match 33.8%; Score 673.8; DB 12; Length 715;
Best Local Similarity 98.0%; Pred. No. 1.4e-116;
Matches 703; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

QY 1134 GGGTGAAGAGGTGGACCAAAAGCTGGACATCTTCAATAGGAGCTACTGCTAATCCCA 1193
|||||
Db 1 GGGTGAAGAGGTGGACCAAAAGCTGGACATCTTCAATAGGAGCTACTGCTAATCCCA 60
|||||

QY 1194 TCACCTGGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTGAGGCGACGACCATCACCT 1253
|||||
Db 61 TCACCTGGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTGAGGCGACGACCATCACCT 120
|||||

QY 1254 ATTTGACTCGAGCGTACCTTAACCGCGCTGCCCTTAAGCATATGCGCAAGTATCTAC 1313
|||||
Db 121 ATTTGACTCGAGCGTACCTTAACCGCGCTGCCCTTAAGCATATGCGCAAGTATCTAC 180
|||||

QY 1314 AGCAGAGCGGTGAAGAGACCGACTGATTTCCACGAGGCTGGAAGGTTACTTCA 1373
|||||
Db 181 AGCAGAGCGGTGAAGAGACCGACTGATTTCCACGAGGCTGGAAGGTTACTTCA 240
|||||

QY 1374 AAATGAATGTGGCAGGCAGAAATATGACAGTGACTGTGGTGTCTTTGTGTCAGTACT 1433
|||||

```

```

Db 241 AAATGAATGTGGCAGGCAGAAATATGACAGTGACTGTGGTGTCTTTGTGTCAGTACT 300
|||||
QY 1434 GCAAGCATCTGCCCTGTCTCAGCCATTTCAGCTTTCACCCAGCAGGACATGCCCAAACTTC 1493
|||||
Db 301 GCAAGCATCTGCCCTGTCTCAGCCATTTCAGCTTTCACCCAGCAGGACATGCCCAAACTTC 360
|||||
QY 1494 GTCCGCAGATCTACAAGAGAGCTGTGTCACTGCAAACTCACTGTGTGAGGCTCGTACCCCA 1553
|||||
Db 361 GTCCGCAGATCTACAAGAGAGCTGTGTCACTGCAAACTCACTGTGTGAGGCTCGTACCCCA 420
|||||
QY 1554 GACCCCAAGCCCAATAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAAATCCAGTT 1613
|||||
Db 421 GACCCCAAGCCCAATAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAAATCCAGTT 480
|||||
QY 1614 CTTTTCCTCTCTGCTCTTCCCACTCACTCCCTTGGTGTGTTTTCATATTTAAATCTTTC 1673
|||||
Db 481 CTTTTCCTCTCTGCTCTTCCCACTCACTCCCTTGGTGTGTTTTCATATTTAAATCTTTC 540
|||||
QY 1674 AATTTCTGATTTTCTTTTGAGAGAACTCTTCTGATTTCTGATGTGCAGGGGGTG 1733
|||||
Db 541 AATTTCTGTA-TTGTGTTTCTTTTGAGAGAACTCTTCTGATGTGCAGGGGGTG 598
|||||
QY 1734 GTTACAGAAAAGCCCTTCTCTCTGTTTGCAGGGGAGTGGCCCTGTGCGCTGGGT 1793
|||||
Db 599 GTTACAGAAAAGCCCTTCTCTCTGTTTGCAGGGGAGTGGCCCTGTGCGCTGGGT 658
|||||
QY 1794 GGAGCAGTCATCTCTCCCTTCCCTGTCAGGAGCAGGAGAAATCAGTGTGGGGTG 1850
|||||
Db 659 GGAGCAGTCATCTCTCCCTTCCCTGTCAGGAGCAGGAGAAATCAGTGTGGGGTG 715
|||||

RESULT 12
AL579669/c
LOCUS AL579669 LTI_FL012_Tc1 Homo sapiens cDNA clone CS0DH007YF21 3 prime
DEFINITION mRNA sequence.
ACCESSION AL579669
VERSION AL579669.1 GI:12944937
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 662)
JOURNAL Li W.B., Gruber, C., Jessee, J. and Polayes, D.
COMMENT Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..662
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DH007YF21"
/clone_lib="LTI_FL012_Tc1"
/tissue_type="T cells from T cell leukemia"
/lab_host="DH10B"
Note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-stranded cDNA was digested with Not I and cloned
into the Not I and Eco RV sites of the pCMVSPORT 6 vector.
Library was constructed by Life Technologies. Contact :
Feng Liang Life Technologies, a division of Invitrogen
9800 Medical Center Drive Rockville, Maryland 20850, USA
Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 167 a 158 c 162 g 147 t 28 others
ORIGIN
Query Match 31.9%; Score 634.2; DB 9; Length 662;

```

Best Local Similarity 94.9%; Pred. No. 3.9e-109;  
Matches 628; Conservative 23; Mismatches 11; Indels 0; Gaps 0;

```
QY 1215 CCCTCATCTCTGTTGATGTGAGGCGCGACCACTACCTATTGTTGACTCGCAGGTACCC 1274
Db 662 CCCTCATCTCTGTTGATGTGAGGCGCGACCACTATTGTTGACTCGCAGGTACCC 603
QY 1275 TAAACCCGCGTCCCTTAAGCATATTCGCAAGTATCTACAGGAGAGCGGTAAAGAAAG 1334
Db 602 TAAACCCGCGTCCCTTAAGCATATTCGCAAGTATCTACAGGAGAGCGGTAAAGAAAG 543
QY 1335 ACCGACTGGATTTCCACAGGCGTGGAAAGGTTACTTTCAAATGAATGTGGCCAGGAGA 1394
Db 542 ACCGACTGGATTTCCACAGGCGTGGAAAGGTTACTTTCAAATGAATGTGGCCAGGAGA 483
QY 1395 ATATGACAGTACTGTGGTGCCTTTGTTGAGTACTCAAGCATCTGGCCCTGTCTC 1454
Db 482 ATATGACAGTACTGTGGTGCCTTTGTTGAGTACTCAAGCATCTGGCCCTGTCTC 423
QY 1455 AGCATTTCAGCTTCACCCAGCAGGACATGCCAAACTTCGTCGGCAGATCTACAAGGAGC 1514
Db 422 AGCATTTCAGCTTCACCCAGCAGGACATGCCAAACTTCGTCGGCAGATCTACAAGGAGC 363
QY 1515 TGTGTCACTGCAAACTCACTGTGTGAGCCCTGTACCCAGACGCCCAAGCCCAATAATGGG 1574
Db 362 TGTGTCACTGCAAACTCACTGTGTGAGCCCTGTACCCAGACGCCCAAGCCCAATAATGGG 303
QY 1575 AAGGAGACATGGAGTCCCTCCCAAGAACTCCAGTTCCTTCTCTTGCCTTTC 1634
Db 302 AAGGAGACATGGAGTCCCTCCCAAGAACTCCAGTTCCTTCTCTTGCCTTTC 243
QY 1635 CCACACTCTCCCTTTGGTGTTCATATTTAAATGTTTCAATTTCTGATTTTCTTCT 1694
Db 242 CCACACTCTCCCTTTGGTGTTCATATTTAAATGTTTCAATTTCTGATTTTCTTCT 183
QY 1695 TTGAGAAATACTTGTGATTTCTGATGTGAGGCGGTGCTACAGAAAGCCCTTCTTCT 1754
Db 182 TTGAGAAATACTTGTGATTTCTGATGTGAGGCGGTGCTACAGAAAGCCCTTCTTCT 123
QY 1755 TCCTCTCTTTGAGGAGGTGGCCCTGCTGGTGGGAGCAGTCATCTCCCTTCT 1814
Db 122 TCCTCTCTTTGAGGAGGTGGCCCTGCTGGTGGGAGCAGTCATCTCCCTTCT 63
QY 1815 CCCGTCAGGAGGACAGAAATCAGTCTGCGGGGTGTTGGGCGGACAAATAGGATCACTGC 1874
Db 62 CCCGTCAGGAGGACAGAAATCAGTCTGCGGGGTGTTGGGCGGACAAATAGGATCACTGC 3
```

RESULT 13  
AW961869  
LOCUS  
DEFINITION  
EST373942 MAGE resequences, MAGG Homo sapiens cdna, mRNA sequence.  
ACCESSION  
AW961869  
VERSION  
AW961869.1 GI:8151555  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 676)  
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gasparid,R., Gay,C., Holt  
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and  
Quackenbush,J.  
TITLE  
Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cdna microarray  
JOURNAL  
Unpublished (2000)  
COMMENT  
Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnqu@tigr.org  
Plate: 167  
Seq primer: Reverse.  
Location/Qualifiers  
source  
1. .676  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGG"  
/note="Vector: pBluescriptskm"

BASE COUNT 135 a 191 c 212 g 138 t  
ORIGIN

Query Match 31.6%; Score 629.4; DB 10; Length 676;  
Best Local Similarity 98.1%; Pred. No. 3.1e-108;  
Matches 658; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

```
QY 334 GGGGTGCGAGCCGACAGGTGCCATCCCTGTTGTTGCTTTTGACTCCCCCGGGGGCCAC 393
Db 2 GGCACGAGGCGCCACAGGTGCCATCCCTGTTGCTTTTGACTCCCCCGGGGGCCAC 61
QY 394 CTCACCCCGGCTGGGTGCTGCTAGTGTCTCATGGCTGAGGATGGGGTGAGAGGGTCTC 453
Db 62 CTCACCCCGGCTGGGTGCTGCTAGTGTCTCATGGCTGAGGATGGGGTGAGAGGGTCTC 121
QY 454 CACCACTGCCCTCTGGGCCCCCATGGAGAACTGAGTCACTCAGTGGACTCCAAAGTCTC 513
Db 122 CACCACTGCCCTCTGGGCCCCCATGGAGAACTGAGTCACTCAGTGGACTCCAAAGTCTC 181
QY 514 CTCGTGACCTGACTCGGGCTCCTTTTCATGTACTCTGCCCAACGGTTTTTGGGGGACAAT 573
Db 182 CTCGTGACCTGACTCGGGCTCCTTTTCATGTACTCTGCCCAACGGTTTTTGGGGGACAAT 241
QY 574 CTGGGCGAGAGGGGAGCGCA-CTTGGCACCCTCTGATGCCACATCCCTCATCAGCAATG 632
Db 242 CTGGGCGAGAGGGGAGCGCACTTGGCACCCTCTGATGCCACATCCCTCATCAGCAATG 301
QY 633 TGTGACAGTCGCGGACCATGTGGCCAGGAGCTTTTTCAGGGCTCAGATTTGGGCGATGG 692
Db 302 TGTGACAGTCGCGGACCATGTGGCCAGGAGCTTTTTCAGGGCTCAGATTTGGGCGATGG 361
QY 693 CAGAAGAGCGAGAGGCGCTGGGGAGAAAGCCGCGCCAGCACAGCCCCCTTCGAGAGGAGC 752
Db 362 CAGAAGAGCGAGAGGCGCTGGGGAGAAAGCCGCGCCAGCACAGCCCCCTTCGAGAGGAGC 421
QY 753 ATGTGACCTGCGTACAGAGCATCTTGGACGAATTCCTTCAACGATGAGCAGCTCATAC 812
Db 422 ATGTGACCTGCGTACAGAGCATCTTGGACGAATTCCTTCAACGATGAGCAGCTCATAC 481
QY 813 CCCTCAGCACTGATGAGGTAGTAGAAGCTGGAGGACATTTTCCAGCAGGAGCTTTTCCA 872
Db 482 CCCTCAGCACTGATGAGGTAGTAGAAGCTGGAGGACATTTTCCAGCAGGAGCTTTTCCA 541
QY 873 CCCCTTCCAGGAAGGCGCTGGTGTTCAGCTGATCAGTCTTTACCAGCGGATCCCAAGGCA 932
Db 542 CCCCTTCCAGGAAGGCGCTGGTGTTCAGCTGATCAGTCTTTACCAGCGGATCCCAAGGCA 601
QY 933 ATGCCATGTTGAGGGCTTCCGAGTGGCTTAT-AAAGCGGACGCTGTGACCATGGATGAC 991
Db 602 ATGCCATGTTGAGGGCTTCCGAGTGGCTTATAAAGCGGACGCTGTGACCATGGATGAC 661
QY 992 TTGGGGACCTT 1002
Db 662 CTGGGGGACCT 672
```

RESULT 14  
BG870520  
LOCUS  
DEFINITION  
602791439F1 NCI\_CGAP\_SG2 Mus musculus cdna clone IMAGE:4922696 5',  
mrna sequence.  
ACCESSION  
BG870520

VERSION BG870520.1 GI:14221060  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 804)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapps-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM10842 row: n column: 09  
High quality sequence stop: 789.  
FEATURES  
Source  
1..804  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4922696"  
/clone\_lib="NCI-CCAP-SG2"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site\_1:  
NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI-CCAP Library."  
BASE COUNT 223 a 195 c 202 g 184 t  
ORIGIN  
Query Match 31.6%; Score 628.4; DB 12; Length 804;  
Best Local Similarity 90.5%; Pred. No. 4.7e-108;  
Matches 693; Conservative 0; Mismatches 71; Indels 2; Gaps 2;  
QY 819 GCACGTGACGTAGTACAGAGCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTT 878  
DB 33 CGGGGATGAGTTGTAGAGAGTTGGAGGACATTTTCCAGCAGGAGTTCTACACCC 92  
QY 879 CCAGGAAGGGCTGTGTGACGTGATCCAGTCTTACAGCGGATGCCAGCAATGCCA 938  
DB 93 CAAGAAGAGTCTGTACTACAGTGTCCAGTCTTATACCGGATGCCAGCAAGCTA 152  
QY 939 TGGTGAGGGGCTCCGAGTGGCTTATAAGCGGCACGCTGTACCATGGATGACTTTGGGA 998  
DB 153 TGGTAAGGGGCTCCGGGTATCCTATAAGCGACACGCTGTACCATGGATGACTTTGGGA 212  
QY 999 CTTTGTATGGACAGAACTGGTCAATGATACCAGGTGATGAACATGTATGAGACCTGGTCA 1058  
DB 213 CTTTATATGGACAGAACTGGTCAATGATACCAGGTGATGAACATGTATGAGATCTGGTCA 272  
QY 1059 TGGACACAGTCCCTGAAAGGTGCATTTCTCAATAGTTTCTTATGATGAACATCCGTA 1118  
DB 273 TGGACACAGTCCCTGAAAGGTGCATTTCTCAACAGCTTCTTATGATGAACATCCGTA 332  
QY 1119 CAAGGGTTATGATGGGTGAA -AAGGTGGACCAAAACGGTGACATCTTCAATAAGGAG 1177  
DB 333 CCAAGGGTTATGATGGGTGAAAGGAGGTGGACCAAAATGTGACATCTTCAATAAGGAA 392  
QY 1178 CTACTGCTAATCCCATCCACCTGAGGTGCAATGGTCCCTCATCTCTGTTGATGTAGG 1237  
DB 393 TTACTGCTAATCCCATCCATCCATGGAGGTGCACCTGGTCCCTTATCTCATGTTGATGAAG 452  
QY 1238 CGACGACCATCACCTATTTTGACTCGCAGCGTACCCCTAAACCGCGCTGCCCTTAAGCAT 1297  
DB 453 CGACGTACCATCACCTATTTTGACGTCGCCAGCGAATCTTAATTCGCGCTGCCCTTAAGCAT 512  
QY 1298 ATTGCCAAGTATCTACAGCAGAGGGGTAAAGAAAGACCGGACTGGATTTCACACAGGCG 1357

DB 513 ATTGCCAAGTATCTACAGCAGAGGCGAGTCAAAAAGACCGGACTTCCATCAGGCG 572  
QY 1358 TGGAAAGGTTACTTCAAAATGAATGTGGCCAGCAGAGAATAATACAGTCACTGTGTGCT 1417  
DB 573 TGGAAAGGTTACTTCAAAATGAATGTGGCCAGCAGAGAATAATACAGTCACTGTGTGCT 632  
QY 1418 TTTGTGTTCAGTACTCAAGCATCTGGCCCTGTCTTCAGCCATTCAGCTTCCACCCAGAG 1477  
DB 633 TTTGTATTACACTACTGCAAGCACCTGGCCCTGTCTCAGCCATTCAGCTTCCACCCAGAG 692  
QY 1478 GACATGCCCAAACTTCGTGGCAGAGTCTACAAGAGAGTCTGTCTACCTGCAAACTCACTGTG 1537  
DB 693 GACATGCCCAAACTTCGTGGCAGAGTCTACAAGAGAGTCTGTCTACCTGCAAACTCACTGTG 752  
QY 1538 -TGACCCCTGCTACCCAGACCCCAAGCCCAATAAATGGGAAGGAGA 1582  
DB 753 TTGAGCCTCATAAACCCAGACCTTAAAGCCATCATGCGGAGGGGAAA 798  
RESULT 15  
LOCUS BM596661  
DEFINITION BM596661  
ACCESSION BM596661  
VERSION BM596661.1 GI:19009919  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 626)  
REFERENCE Bonaldo, M.F., Lennon, G. and Soares, M.B.  
AUTHORS Normalization and subtraction: two approaches to facilitate gene  
TITLE discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 9704477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics ([www.resgen.com](http://www.resgen.com)).  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1..626  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-E-DW0-agk-m-22-0-UI"  
/clone\_lib="UI-E-DW0"  
/tissue\_type="lens"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a  
modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
UI-E-DW0 is a cDNA library containing the following  
tissue(s): lens. The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was ligated to an EcoR I adaptor, digested with Not  
I, and cloned directionally into pT7T3-Pac vector. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The

sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT	159 a	157 c	172 g	138 t	
ORIGIN					
Query Match	31.4%	Score 626;	DB 14;	Length 626;	
Best Local Similarity	100.0%	Pred. No. 1.3e-107;			
Matches 626;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY 694	AGAGAGGCGAGAGGGCTGGGAGAGAAAGCCGGCGAGCAGAGCCCGCTGGAGAGAGCA	753			
Db	1 AGAAGAGGCGAGAGGGCTGGGAGAGAAAGCCGGCGAGCAGAGCCCGCTGGAGAGAGCA	60			
QY 754	TGTGACCTGGCTACAGAGCATCTTGGAGCAATTCCTTCAACGCTATGGCAGCCTCATACC	813			
Db	61 TGTGACCTGGCTACAGAGCATCTTGGAGCAATTCCTTCAACGCTATGGCAGCCTCATACC	120			
QY 814	CCTCAGCACTGATGAGTAGTAGAGAGCTGGAGGACATTTCCAGCAGGAGTTTCCAC	873			
Db	121 CCTCAGCACTGATGAGTAGTAGAGAGCTGGAGGACATTTCCAGCAGGAGTTTCCAC	180			
QY 874	CCCTTCCAGGAAGGGCTGGTGTGGAGCTGATCCAGTCTTACCGCGGATGCCAGGCAA	933			
Db	181 CCCTTCCAGGAAGGGCTGGTGTGGAGCTGATCCAGTCTTACCGCGGATGCCAGGCAA	240			
QY 934	TGCCATGGTCAGGGGCTTCGGAGTGGCTTATAAGCGGACGTGCTGACCATGGATGACTT	993			
Db	241 TGCCATGGTCAGGGGCTTCGGAGTGGCTTATAAGCGGACGTGCTGACCATGGATGACTT	300			
QY 994	GGGGACCTTCTATGGACAGAACTGGGTCAATGACAGGTGATGAACATGATGGAGACCT	1053			
Db	301 GGGGACCTTCTATGGACAGAACTGGGTCAATGACAGGTGATGAACATGATGGAGACCT	360			
QY 1054	GGTCATGGACACAGTCCCTGAAAAGGTGCATTTCTTCAATAGTTTCTTCTATGATAACT	1113			
Db	361 GGTCATGGACACAGTCCCTGAAAAGGTGCATTTCTTCAATAGTTTCTTCTATGATAACT	420			
QY 1114	CCGTACCAAGGGTTATGATGGGTGAAAAGGTGGACCAAAAACGTGGACATCTTCAATAA	1173			
Db	421 CCGTACCAAGGGTTATGATGGGTGAAAAGGTGGACCAAAAACGTGGACATCTTCAATAA	480			
QY 1174	GGAGCTACTGCTAATCCCCATCCACCTGGAGGTGCATTTGGTCCCTCATCTCTGTTGATGT	1233			
Db	481 GGAGCTACTGCTAATCCCCATCCACCTGGAGGTGCATTTGGTCCCTCATCTCTGTTGATGT	540			
QY 1234	GAGGCGACGACCATCACCTATTTTGTACTCGGAGGTACCCCTAAACGGCGCTGCCCTAA	1293			
Db	541 GAGGCGACGACCATCACCTATTTTGTACTCGGAGGTACCCCTAAACGGCGCTGCCCTAA	600			
QY 1294	GCATATTGCCAAGTATCTACAGGCAG	1319			
Db	601 GCATATTGCCAAGTATCTACAGGCAG	626			

Search completed: December 16, 2002, 17:45:56  
Job time : 2574 secs